



**What You Don't Know Won't Hurt You:  
A Laboratory Analysis of Betrayal Aversion**

**by**

**Jason A. Aimone & Daniel Houser**

**Interdisciplinary Center for Economic Science  
George Mason University**

**September 2008**

# **What You Don't Know Won't Hurt You: A Laboratory Analysis of Betrayal Aversion**

**Jason A. Aimone & Daniel Houser**

**September 2008**

**Interdisciplinary Center for Economic Science  
George Mason University**

**Abstract:** Trust promotes economic growth and development, and previous research has shed much light on reciprocity and other motives for trusting decisions. Why people choose not to trust has received substantially less attention, perhaps in part because not trusting is predicted by standard economic theory: selfish people consider the (perhaps subjective) stochastic nature of the environment and make the earnings-maximizing decision. This explanation is incomplete: we provide evidence from a laboratory analysis with an investment game that people's decisions vary according to how an environment's uncertainty will be resolved. In particular, if resolving uncertainty requires an investor to learn whether her trustee chose to betray then she is much less likely to trust. Our data thus provide evidence that "betrayal aversion" detrimentally affects propensities for trusting decisions. Our results also emphasize the importance of impersonal, institution-mediated exchange in promoting investment and economic efficiency.

**Acknowledgements:** We thank Ernst Fehr and Benjamin Schoefer for a valuable discussion that led to the experiment design reported in this paper. For helpful comments and suggestions, we thank Vernon Smith, Bart Wilson, Erte Xiao and participants in research seminars at ICES, the Mercatus Center, George Mason University, University of Zurich, and the International ESA 2008. Kail Padgitt and Adam Smith provided capable laboratory assistance. We gratefully acknowledge research support from the International Foundation for Research in Experimental Economics.

## 1 Introduction

First movers in bargaining environments often must decide whether to forgo guaranteed returns in order to trust counterparts to provide them with greater future benefits (e.g., efficiency wages in labor markets; see Fehr and Falk 1999; Rigdon, 2002.) While both naturally occurring and designed environments have established that many people do trust, some people of course do not (see, e.g., Berg et al. 1995.) Choosing not to trust is consistent with standard economic theory, and perhaps for this reason few alternative explanations for such decisions have appeared. This is unfortunate both because choosing not to trust might involve decision processes distinct from those that standard theory suggests, and because designing institutions to promote trust requires one to understand why sometimes people choose not to do so.

To explain the decision not to trust, standard theory posits that selfish people consider the stochastic nature of their environment, forming subjective estimates of the likelihood of reciprocity, and then make the expected earnings-maximizing decision. We here provide evidence from a laboratory analysis of an investment game that this explanation is incomplete. In particular, we find that people's trusting decisions vary according to how an environment's uncertainty is resolved. If resolving uncertainty requires an investor to learn whether her trustee chose to betray her then she is much less likely to trust. Accordingly, our data provide evidence that "betrayal aversion" detrimentally affects propensities for trusting decisions, thus emphasizing the importance of impersonal, institution-mediated exchange in promoting trust, investment and efficient economic outcomes.

Recent interest in effects of betrayal aversion on trust decisions owes largely to research by Bohnet and Zeckhauser (2004); Bohnet et al. (2006); and Bohnet et al. (2008). Those contributions are important first steps in shedding light on betrayal aversion and its effects on economic exchange behavior. Each of those papers used a Becker-DeGroot-Marschak (1964) procedure to elicit the "minimum acceptable probability" (MAP) of positive returns one requires to play games in environments of strategic or state uncertainty. (The former refers to uncertainty related to a counterpart's decision in a strategic environment, and the latter to uncertainty related to non-strategic factors not controlled by another person, such as the outcome of a roll of a die.) They find

that elicited probabilities (MAPs) systematically differ between environments in a way that is consistent with betrayal aversion. We discuss further below, however, that factors distinct from betrayal can also plausibly lead to the differences they discovered.

Here we report data from one-shot two-person binary investment games (Tullock, 1967<sup>1</sup>) in which investors can choose not to know the decision of their particular trustee, and instead receive payment according to a random draw from a separate pool of decisions identical to the pool of trustees' decisions. Note that the probability of receiving the "cooperative" outcome is identical in the two cases, and participants understand this is the case. Our design differs from the MAP-based design primarily in that it does not require one to elicit probabilities: our inferences stem from revealed-preferences. We argue below that this is a significant improvement in the sense that it allows us to avoid certain confounds that make the interpretation of MAP-based designs difficult.

Our main finding is that investors systematically prefer to remain ignorant of their specific trustee's decision. Moreover, when avoiding this information is not possible investors are substantially less likely to make trusting decisions. These results are convergent evidence that outcome-based models cannot fully explain economic decision making in strategic environments (see, e.g., McCabe, et al. 2003.) More specifically, our findings suggest that impersonal institution-mediated exchange (e.g., lending through financial intermediaries) promotes trusting decisions and economic efficiency by shielding investors from knowing whether their particular trustee chose to betray.

Our data also reinforce the general importance of emotions to economic decision-making (see, e.g., Fehr et al. 2005; Xiao and Houser, 2005), and provide new evidence on the specific importance of emotion regulation (see, e.g., Gross, 1998.)<sup>2</sup> Moreover, our design suggests a way to construct institutional solutions to inefficiencies stemming from betrayal aversion, an important topic that previous investigations have not addressed.

Finally, it is worth noting that our investigation contributes to the literature on distinctions between trust and risk environments (see, e.g., Houser et al., 2008; Schechter, 2007; Kosfeld et al., 2005; Eckel and Wilson, 2004; Ashraf et al. 2006; McCabe et al.,

---

<sup>1</sup> Tullock (1967) is the first paper, to our knowledge, to discuss and work through the implications of a sequential prisoner's dilemma within a trust context.

<sup>2</sup> In our case, emotion regulation can explain the actions people take to avoid the negative emotional experience of betrayal.

2001; Snijders and Keren, 1998.) This literature provides substantial evidence that people make trusting decisions differently than decisions under risk. However, identifying the source of the differences is difficult. A reason is that trust and risk environments typically differ in multiple ways (e.g., strategic-uncertainty always involves another person, while state-uncertainty need not.) As a result, while it is widely accepted that trusting decisions differ between environments of strategic and state-uncertainty, the reason for such differences – and particularly the role of betrayal aversion – remains an important open question.

The following section reviews the relevant betrayal literature. Our experiment design is in section 3. Section 4 gives predictions and hypotheses, and section 5 describes our data and results. Section 6 is a concluding discussion.

## **2 Background**

A “betrayal” is an intentional action negatively inconsistent with shared expectations encompassing a relationship of mutual trust (Elangovan and Shapiro, 1998.) Such actions would include, for example, the betrayals by Judas Iscariot, Brutus or Benedict Arnold. To experience betrayal is exceptionally emotionally costly. As noted by Jackson (2003), “the betrayed experience powerful sensations of violation; they feel used and damaged”<sup>3</sup> (see also Finkel et al., 2002.) Humans respond strongly to the thought of betrayal, even when this betrayal does not involve them. For example, survey evidence gathered by Koehler and Gershocoff (2003) indicates that people punish hypothetical crimes including an aspect of betrayal much more severely than similar hypothetical crimes lacking this element.

Humans surely prefer not to experience betrayal and, consequently, might prefer to avoid situations where betrayal might occur. Such avoidance decisions are consistent with the notion of “emotion-regulation” which has emerged in recent psychology and neuroscience literature. Studies in this area (see, e.g., surveys by Gross (1998) and Ochsner and Gross (2005)) argue that emotion-regulation, at both the conscious and subconscious level, is an important reason that expectations affect decisions. For example, emotion regulation can explain why people often choose to turn conversations

---

<sup>3</sup> Jackson, p. 72.

away from topics they expect could become uncomfortable. More generally, emotion-regulation can underlie any decision not to take an action that one expects might lead to unpleasant emotional outcomes.

Bohnet and Zeckhauser (2004), Hong and Bohnet (2007) and Bohnet et al. (2008) each investigate the role of betrayal aversion in economic decision-making. Each of these papers uses the same experiment design. In brief, the investigators asked each subject to report the “minimum acceptable probability” (MAP) at which s/he would choose a (trust or risk) “gamble.” Consistent with betrayal aversion and the previous research on the distinction between risk and trust games (Houser et al. 2008, Eckel and Wilson 2004, etc.), these papers found subjects report higher average MAPs for trust gambles than risk gambles. However, as noted by Bohnet and Zeckhauser (2004), several other factors unrelated to betrayal could be at work in generating differences in behavior between treatments.<sup>4</sup> Indeed, Bohnet et al. (2006) discuss data from this same design within a loss aversion framework that does not appeal to betrayal considerations. Whether or not betrayal aversion drives these results is currently unknown.

Previous studies have also considered the effect of individual differences, and especially gender, in betrayal aversion. Bohnet and Zeckhauser (2004) find evidence of substantial betrayal aversion in both genders, while Hong and Bohnet (2007) seem to find little evidence that women are betrayal averse. The three survey studies of forgiveness after betrayal in Finkel et al. (2002) come to three different conclusions: one finds men are more likely to forgive betrayal than women; another suggests less likely; and a third finds no difference. Finally, in a nice survey paper, Croson and Gneezy (2008) argue the balance of the evidence suggests women seem less willing to trust than men. This might suggest that women exhibit greater betrayal aversion than men.

Finally, the possibility of betrayal underlies much of the reputations literature (e.g., Houser and Wooders, 2006), and has also received attention within neuroeconomics. A notable example is Kosfeld et al. (2005), which draws a connection

---

<sup>4</sup> Bohnet and Zeckhauser (2004) point to disutility from loss of control, assessment costs associated with calculating trustworthiness, costs of making incorrect assessments, costs from placing trustees in a potentially undesirable decision situation, and disutility from earning money due to other people’s kindness as factors that could lead to differences between treatments.

between oxytocin and trust decisions, arguing that oxytocin helps humans to overcome trust obstacles including betrayal.

Our research is novel in that it provides the first compelling evidence that investors prefer to avoid knowing whether a trustee choose to betray them, and when such knowledge is unavoidable they are much more likely not to make trusting decisions. The next section details the experiment design we used to obtain these conclusions.

### **3 Experiment Design**

Our goal is to create an experiment that rigorously identifies the effect of betrayal aversion on propensities to form trust relationships in economic exchange environments. To do this we consider three treatments of a one-shot binary trust game (see Figure 1). In our game a human investor in room A (player 1) and a human trustee in room B (player 2) make decisions simultaneously.<sup>5</sup> The way investors make decisions varies by treatment, as described below. Appendix A provides a transcript of the instructions for all treatments.

<Figure 1>

#### **3.1. KNOW Treatment**

In the “KNOW” treatment each investor chooses either “trust” or “don’t trust”, and the trustee simultaneously chooses either “betray” or “reciprocate.”<sup>6</sup> If the investor chooses “don’t trust” then both players earn five dollars regardless of the trustee’s decision. If the investor chooses “trust” then payoffs depend on the trustee’s decision. In particular, if the trustee chooses to reciprocate then each subject receives 15 dollars, while if the trustee chooses to betray then the investor receives two dollars and the trustee receives 28 dollars.

---

<sup>5</sup> Note that this game tree was not distributed to subjects, nor were the terms “trust”, “betray”, or “reciprocate”, used to describe the game.

<sup>6</sup> The instructions also contain neutral framing. In particular, the words “trust”, “betray” and so on did not appear.

### 3.2. OPTION-TO-KNOW Treatment

The second treatment, the “OPTION-TO-KNOW” (OPTION) treatment, is the same as KNOW except each investor has the additional option to be paid according to a computer’s draw from the pool of that specific session’s trustees’ decisions. The instructions inform investors that they have both a randomly assigned human counterpart as well as a randomly assigned computer-generated “decision” in the experiment. They are assigned to their computer “decision” at the same time as they are assigned to their human counterpart (see procedures below.) An investor in this treatment has three alternatives: (i) “don’t trust”, and thus assign five dollars to both one’s counterpart and oneself; (ii) “trust” and have one’s counterpart’s decision determine one’s payoff; (iii) “trust” and have a computer generated “betray” or “reciprocate” decision determine one’s payoff. Note that in the second and third cases the investor is aware that her trustee receives payment according to the trustee’s own decision. Consequently, a trustee’s earnings are identical regardless whether the investor chose payment based on the human counterpart’s decision or the computer’s decision, and are independent of the computer’s decision.

The third treatment, the “DONTKNOW” treatment, is the same as the KNOW and OPTION treatments except the investors choose between two alternatives: the “don’t trust” option (the \$5/\$5 split) and the computer option described above<sup>7</sup>. Investors’ human trustees make the same decision as in the KNOW and OPTION treatments, and are paid in the same manner as well.

The instructions (truthfully) inform subjects that the computer makes its decision as follows. First, a computer tallies the number of “betray” and “reciprocate” decisions the N trustees made in that particular session. Next, the computer randomly assigns the numbers one through N either “betray” or “reciprocate.” Thus, there are exactly the same number of “betray” and “reciprocate” decisions in the N computer assignments as in the trustees’ data. It follows that the probability a trusting investor receives a “reciprocate” decision from her (randomly assigned) trustee is identical to the chance that the investor

---

<sup>7</sup> As the room B instructions say that a possible scenario is for the trustee’s decision to determine both subjects’ payoffs, in DONTKNOW sessions one of the room A subjects was given a KNOW or OPTION set of instructions in a different room. For this reason, we were able to use identical room B instructions in all treatments. The decisions from the few separated subjects have not been analyzed because their environment is not comparable to that of the other OPTION and KNOW treatments’ participants.



receives a (randomly assigned) “reciprocate” decision from the computer. The computer decision can be thought of as a different draw, without replacement, from an identical pool of outcomes, with the exception that there is no *specific* trustee associated with that outcome.

### **3.3. Discussion of Design**

It is worth emphasizing several features of our design. First, trustees’ instructions are identical among treatments, and trustees were in all cases drawn from the same population using identical recruiting procedures. Thus, we expect trustee decisions not to differ systematically between treatments. Also, in an effort to ensure investors’ subjective expectations about trustees’ decisions displayed no systematic variation across treatments, we always gave trustees’ instructions to investors. Controlling subjective expectations has the twin advantages that (i) investors’ strategic uncertainty should not vary systematically among treatments; and (ii) subjects’ reference points are not likely to differ systematically across treatments. Thus, loss aversion is not a plausible explanation for systematic differences in investor decisions among treatments.

It is also important to emphasize that trustees always earn what they choose if the investor chooses “trust”, regardless of whether the investor chooses the human or computer “trust” option. Thus, in relation to choosing not to trust, expected aggregate earnings increase when the investor chooses “trust.” However, the expected increase is invariant to which “trust” option the investor chooses. Because our design ensures subjective investor expectations are invariant among treatments, it follows that the only systematic difference among our conditions is whether one can avoid knowing that one’s specific trustee chose to betray.

### **3.4. Procedures**

Upon arriving to the experiment subjects check in and proceed, as directed, to one of two rooms, room A (for investors) or room B (for trustees.) An equal number of subjects, say  $N$ , are seated in each room. (In our sessions, either eight or ten subjects were in each room.) Once seated, subjects read instructions and the experimenter reads the instructions aloud. Participants then answer a short quiz to ensure they understand the

environment. After all participants successfully complete the quiz, each investor draws a number, without replacement, from a box containing the integers one through N. The outcome of that draw pairs the investor with a trustee previously randomly assigned that number.<sup>8</sup> In conditions with a computer decision option, the same number drawn pairs them to one of the N randomly generated computer decisions. After this, the experiment proceeds according to the treatment conditions. After privately receiving their results, subjects respond to a general questionnaire. Subjects receive payment in cash, privately, immediately prior to leaving the laboratory.

#### **4 Hypotheses**

Our main hypothesis that people are betrayal averse places four clear a priori restrictions on data from our treatments. These are as follows.

**Hypothesis 1:** *More investors will choose not to trust in the KNOW condition than the DONTKNOW condition.*

Standard economic theory based on selfish preferences and risk-aversion (but not betrayal aversion) predicts an equal fraction of investors fail to trust in both KNOW and DONTKNOW. The DONTKNOW treatment gives a baseline investor willingness to accept the pure gamble when betrayal is not a consideration (an environment of state-uncertainty.) Our hypothesis that people are betrayal averse implies that adding the possibility of betrayal in the KNOW treatment (an environment of strategic-uncertainty) reduces investors' willingness to trust in relation to DONTKNOW.

**Hypothesis 2:** *The frequency of trust decisions will be at least as high in the combined OPTION-TO-KNOW treatment as found in the higher of the KNOW or DONTKNOW treatments.*

Because Hypothesis 1 is that trust in DONTKNOW should be higher than found in KNOW, we can restate this hypothesis as the frequency of trust in the combined OPTION treatment is expected to be at least as great as found in DONTKNOW. The reason is that one can always choose to play the DONTKNOW game in the OPTION

---

<sup>8</sup> Based on where they sit in the lab.

treatment, and similarly KNOW is available by simply ignoring the other alternative. Thus, the frequency of trust in OPTION should not be lower than is found in the greater of the other two treatments.

**Hypothesis 3:** *The frequency of trust in the KNOW condition will be at least as great as the frequency of trust with humans in the OPTION-TO-KNOW condition.*

If people are on average betrayal averse then they should, on average, prefer to have the computer determine their return instead of a specific human. Note that betrayal aversion does not require all trusting investors to choose the “computer” option in the OPTION treatment. The reason is that some investors in the OPTION treatment might not exhibit betrayal aversion and thus can be indifferent between human and computer alternatives.<sup>9</sup>

**Hypothesis 4:** *Economic efficiency (measured as average earnings) will be lowest in the KNOW condition.*

This hypothesis follows directly from the fact that in our environment investment has a positive return. As a result, environments with greater trust (more investment) will be associated with greater aggregate earnings. Betrayal aversion should tend to diminish trust decisions in the KNOW treatment, in relation to the other treatments, and thus reduce average earnings.

## 5 Results

The experiments took place at the Interdisciplinary Center for Economic Science (ICES) at George Mason University. The randomly recruited subjects from the George Mason student body had no experience with trust games. In addition to any amount earned in the experiment, each subject received seven dollars for arriving to the laboratory on time. Subjects spent about 40 minutes in the laboratory.

---

<sup>9</sup> People might gain a positive utility from satisfying curiosity or the knowledge that their counterpart reciprocated trust, both of which would lead to increased preference for the human trust option.

We report data from a total of 154 subjects in 77 investor-trustee pairs<sup>10</sup>, with 26 pairs in each of KNOW and OPTION, and 25 pairs in DONTKNOW. Figure 2 describes the behavior of trustees by treatment. The fraction of trustees choosing to betray is 67.9%, 69.2%, 61.5% in the DONTKNOW, KNOW and OPTION treatments, respectively. Mann-Whitney tests reveal no significant difference in percentage of subjects choosing to betray. Overall, 66.3% of trustees chose to betray, implying an empirical expected value to trust of \$6.39 as compared to \$5.00 for choosing not to trust.

<Figure 2>

**Result 1:** *Significantly more investors chose not to trust in the KNOW treatment than chose not to trust in the DONTKNOW treatment ( $p < 0.03$ )*<sup>11</sup>.

This result supports Hypothesis 1. Figure 3 describes the behavior of investors in the three treatments. Unlike trustees, significant differences in investor behavior between treatments are apparent. In KNOW, where trusting requires an investor to learn whether their counterpart chose to betray, we observed 65.38% of investors choosing trust. In DONTKNOW, where an investor cannot know whether her counterpart chose to betray, 92% of investors chose to trust. This difference is statistically significant ( $p < 0.03$ ).

<Figure 3>

**Result 2:** *Investors trust significantly more in the OPTION-TO-KNOW treatment than in the KNOW treatment ( $p = 0.001$ ).*

This result supports Hypothesis 2. In OPTION, where an investor can choose to avoid knowing whether she was betrayed, we found that 100% of investors chose to trust. This is significantly greater than the 65.38% who trusted in KNOW ( $p = 0.001$ ). Note that the fraction of investors who chose the trust gamble in DONTKNOW, 92%, does not statistically differ from the 100% who did so in OPTION ( $p = 0.14$ ).

---

<sup>10</sup> As noted in fn 5, the single “separated” investor in each DONTKNOW session (the one who participated in the KNOW or OPTION game) is excluded from our analysis. Our analysis does not exclude any trustee because all trustees in all treatments were in the same situation.

<sup>11</sup> All of the p-values reported in this section are from two-sided Mann-Whitney tests.

**Result 3:** *Significantly more investors chose to trust their human counterpart in the KNOW treatment than in the OPTION-TO-KNOW treatment ( $p = 0.0003$ ).*

This result supports Hypothesis 3. Compared to the 65.38% of investors who chose to trust their human counterpart in KNOW, in OPTION we observe that 46% of investors chose to trust the counterpart while 54% chose the computer trust option (see Figure 4).

<Figure 4>

**Result 4:** *Economic efficiency (measured as average earnings) is lowest in the KNOW treatment.*

This result supports Hypothesis 4. Empirically, we found that the mean earnings among pairs<sup>12</sup> of subjects to be \$30 in OPTION and \$28.06 in DONTKNOW, but only \$23.64 in KNOW. The difference between KNOW and DONTKNOW is statistically significant ( $p < 0.08$ ), as is the difference between KNOW and OPTION ( $p < 0.02$ ). Thus, economic efficiency falls when the possibility of learning of a betrayal is unavoidable.

**Result 5:** *In the KNOW treatment, pairs with female investors earn significantly less than pairs with male investors.*

Of the 77 investors who participated in our experiment, 45 were male and 32 female. In KNOW, mean earnings among pairs with a male investor is \$27.55 ( $n=11$ ), an amount significantly greater than the \$20.78 average ( $n=15$ ) among pairs with female investors ( $p=0.05$ ). That we obtain statistical significance with our rather small sample may suggest that women exhibit greater betrayal aversion than men in this game. We find no significant gender effects on earnings within the OPTION or DONTKNOW treatments. Also, we find no evidence of gender effects in trustees' decisions to betray.

---

<sup>12</sup> Note that if an investor trusts then the expected earnings of a pair of subjects is always 30 dollars regardless of the treatment. If an investor chooses a computer trust option the realized earnings of the pair could be either 17, 30, or 43 dollars, while the expected earnings from trust remains 30 dollars. We use expected earnings from trusting in a particular session for the earnings of an investor in order to have accurate reporting.

## 6 Discussion

We presented, to our knowledge, the first rigorous evidence that failures to trust can be traced to betrayal aversion as distinct from risk-aversion or other factors that appear in standard economic theory based on selfish expected earnings maximization. Our results indicate that less than half of investors choose to know their trustee's decisions when they can avoid such knowledge.<sup>13</sup> Moreover, when such knowledge is unavoidable, a significant fraction simply opt-out by choosing not to invest. Thus, betrayal aversion has a statistically and economically significant detrimental impact on economic efficiency.

Our experimental environment bears some similarity to online anonymous exchange environments (such as eBay, Amazon, etc.) where the identity of trading partners is either anonymous or limited to email addresses or a "store" name. Just as people are averse to betrayal in our experiments, it seems reasonable to expect that people would be averse to betrayal in these online markets as well. Past studies on reputation effects (e.g., Livingston, 2005; Houser and Wooders, 2006) demonstrate that sellers with better reputations earn higher prices. While previous studies have explained this result within the context of state-uncertainty, our findings point to the potential importance of strategic-uncertainty in this environment, and suggest an alternative explanation for reputation premiums: their value lies not only in reducing state-uncertainty but also in reducing the chance of experiencing a negative emotion.

Similarly, a natural explanation for an investor's preference not to know her specific trustee's decision is that she is avoiding the possibility of experiencing a negative emotional outcome. In the psychology literature this is referred to as a "situational modification," and is part of so-called "antecedent-focused emotion regulation" (Gross 1998.) A more complete understanding of how expectations mediate economic decision-making will emerge from an investigation of this area.

---

<sup>13</sup> The questionnaire responses from investors to the question "How would you feel if your counterpart chose D?", seem to support our view that subjects' attitudes are consistent with betrayal aversion. Investors reported they would feel "angry", "miffed", "annoyed", "sad" or "betrayed" if their trust was not reciprocated. One subject, who chose the computer option in OPTION, said betrayal by a human would leave him feeling, "[o]ffended, thus I didn't choose that option." On the other hand, a subject who "trusted" in DONTKNOW replied that if he did not receive the higher payoff, "I would feel neutral because it really is the computer which decides what letter I'm assigned," and another indicated they would feel "Nothing, as my earnings are decided by computer."

Evolution has endowed people with a natural tendency to approach social exchange situations with caution. However, in view of the results of Kosfeld et al. (2005), we speculate that Oxytocin might have co-evolved with betrayal aversion in order to amplify humans' willingness to trust in intimate social relationships. Such co-evolution would help to explain why trust can emerge in contexts where betrayal is especially painful (e.g., trust placed in a dear friend), and yet can fail to develop in less intimate economic-exchange environments. We look forward to exploring this and related hypotheses in future research.

## **References**

- Ashraf, N., I. Bohnet, and N. Piankov (2006): "Decomposing Trust and Trustworthiness." *Experimental Economics* 9:193–208.
- Bohnet, I., F. Grief, B. Herrmann, and R. Zeckhauser (2008): Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *The American Economic Review* 98(1), 294-310.
- Bohnet, I., B. Hermann, and R. Zeckhauser (2006): The Requirements for Trust in Gulf and Western Countries. Working paper, Kennedy School of Government, Harvard University *Forthcoming*.
- Bohnet, I. and R. Zeckhauser (2004): Trust, Risk and Betrayal. *Journal of Economic Behavior and Organization*, 55, 467-484.
- Charness, G., and M. Dufwenberg (2006): Promises and Partnership. *Econometrica*, 74(6), 1579-1601.
- Croson, R. and Gneezy, U. (2008): Gender Differences in Preferences. *Journal of Economic Literature*, *Forthcoming*.
- Eckel, C. and R. Wilson (2004): Is Trust a Risky Decision? *Journal of Economic Behavior and Organization*, 55, 447–465.
- Elangovan, A. R., and D. Shapiro (1998): Betrayal of Trust in Organizations. *Academy of Management Review*, 23, 547-566.
- Fehr, E., U. Fischbacher, and M. Kosfeld (2005): Neuroeconomic Foundations of Trust and Social Preferences: Initial Evidence *American Economic Review*, 95, 346–351.
- Fehr, E. and A. Falk. (1999): Wage Rigidity in a Competitive Incomplete Contract Market. *Journal of Political Economy*. 107, 106-134.
- Finkel, E. J., C.E. Rusbult, M. Kumashiro, and P. A. Hannon (2002): Dealing with Betrayal in Close Relationships: Does Commitment Promote Forgiveness? *Journal of Personality and Social Psychology*, 82, 956-974.
- Gross, J. (1998): The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology*, 2(3) 271-299.
- Holt, C. and S. Laury (2002): Risk Aversion and Incentive Effects in Lottery Choices. *American Economic Review*, 92, 1644-1655.
- Hong, K. and I. Bohnet (2007): Status and Distrust: The Relevance of Inequality and



- Betrayal Aversion. *Journal of Economic Psychology*, 28, 197-213.
- Houser, D., D. Schunk, and J. Winter (2008): Trust Games Measure Trust. Working paper.
- Houser, D., and J. Wooders (2006): Reputation in Auctions: Theory, and Evidence from eBay, *Journal of Economics & Management Strategy*, 15(2), Summer 2006, 353–369.
- Jackson, R. L. (2000): The Sense and Sensibility of Betrayal: Discovering the Meaning of Treachery through Jane Austen. *Humanitas*, 13, 72-89.
- Koehler, J. J., and A. D. Gershoff (2003): Betrayal Aversion: When Agents of Protection Become Agents Of Harm. *Organizational Behavior and Human Decision Processes*, 90, 244-261.
- Kosfeld, M., M. Heinrichs, P. J. Zak, U. Fischbacher, and E. Fehr (2005): Oxytocin Increases Trust in Humans. *Nature*, 435, 673-676.
- Livingston, J. A. (2005): How Valuable is a Good Reputation? A Sample Selection Model of Internet Auctions. *The Review of Economics and Statistics*, 87(3): 453-465.
- McCabe, K., M. Rigdon and V. Smith. (2003): Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization*. 267-275.
- McCabe, K., D. Houser, L. Ryan, V. Smith, and T. Trouard (2001): A Functional Imaging Study of Cooperation in Two-person Reciprocal Exchange. *Proceedings of the National Academy of Science*, 98, 11832-11835.
- Ochsner, K.N., J.J. Gross (2005): The Cognitive Control of Emotion. *TRENDS in Cognitive Sciences*, 9(5) 242-249.
- Rigdon, M. (2002): Efficiency Wages in an Experimental Labor Market. *Proceedings of The National Academy of Science*.
- Schechter, L. (2007): Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay. *Journal of Economic Behavior and Organization*, 62, 272-292.
- Smith, V. L. (1994): Economics in the Laboratory. *Journal of Economic Perspectives*, 8, 113-131.

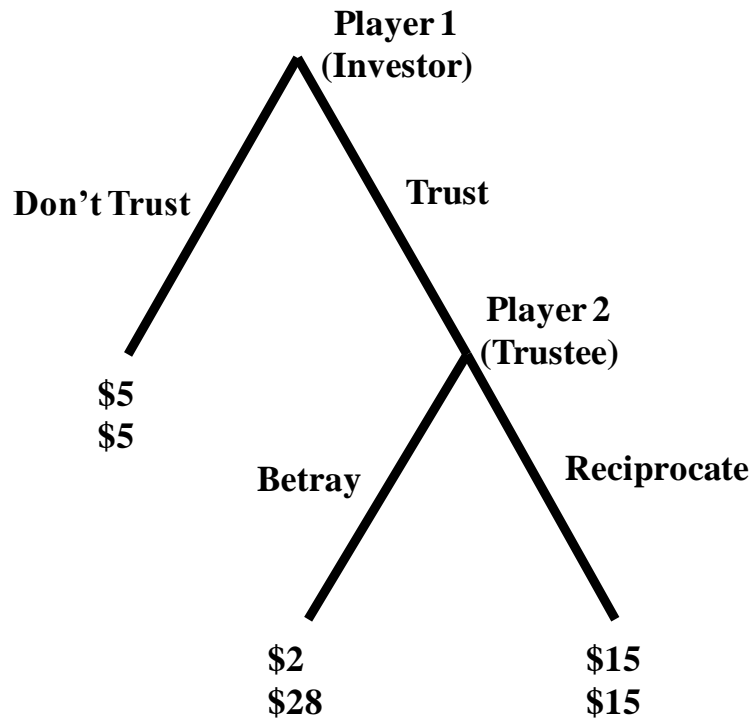
Snijders, C., and G. Keren, (1998) Determinants of trust In: D.V. Budescu, I. Erev and R. Zwick, Editors, *Games and Human Behavior: Essays in Honor of Amnon Rapoport*, Lawrence Erlbaum, Mahwah, NJ (1998), pp. 355–385.

Tullock, G. (1967) The Prisoner's Dilemma and Mutual Trust. *Ethics*, 77, 229.

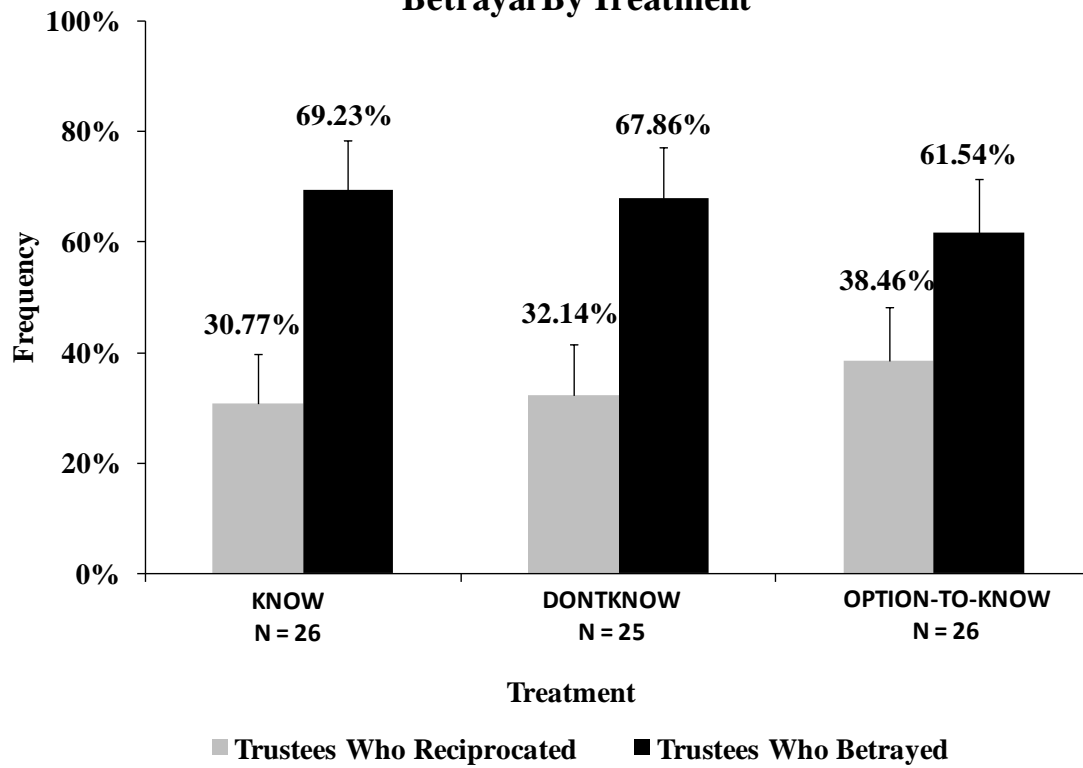
Xiao, E and D. Houser (2005): Emotion Expression in Human Punishment Behavior, *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), 7398-7401.

Vanberg (2008): Why do people keep their promises? An experimental test of two explanations. *Econometrica* (forthcoming.)

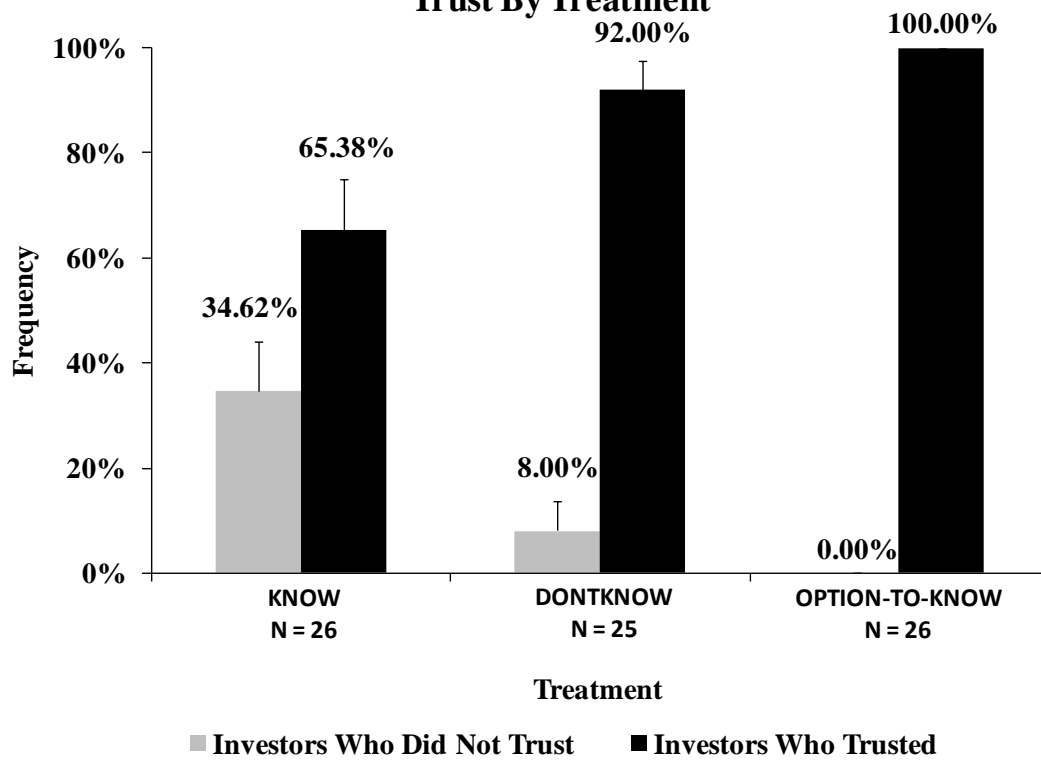
**Figure 1**  
**Investment Game**



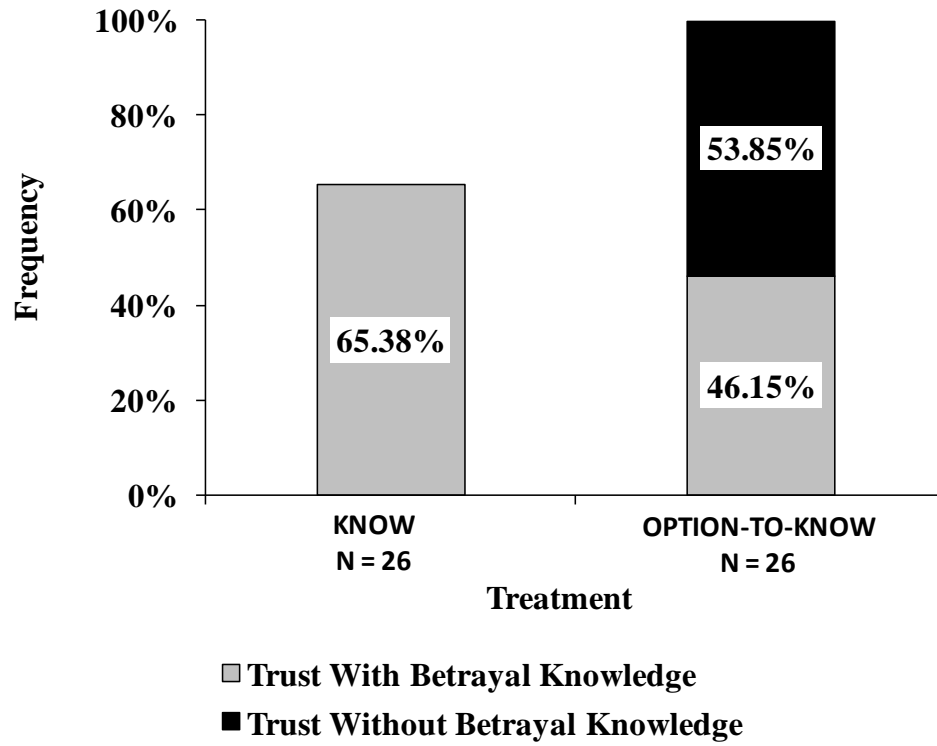
**Figure 2**  
**Betrayal By Treatment**



**Figure 3**  
**Trust By Treatment**



**Figure 4**  
**Type of Trust Decision**  
**By Treatment**



**Appendix A**  
**Room A (Investors) Instructions: OPTION-TO-KNOW Treatment<sup>14</sup>**

Thank you for participating in today's experiment. You've earned a \$7 show-up bonus for participating. In reading and following the instructions below, you have the potential to earn significantly more. You have been randomly assigned to **Room A**. You will also be randomly and anonymously assigned to a person in **Room B**. Your counterpart will not be told your name, and you will not be told his/her name.

**How you are matched with your counterpart:**

Each of the 10 Room A persons will be matched with a different Room B counterpart for the entire experiment. The experimenter will bring around a box with the numbers 1 through 10 inside. The number you draw will assign you to one of the 10 counterparts in Room B (B1 through B10 coinciding with the numbers 1 through 10 in the box). The number also matches you with one of the 10 computer number decisions (coinciding with numbers 1 through 10 in the box).

**Your Decision:**

You have three options for how the earnings for you and your counterpart will be determined in today's experiment. You must choose exactly one of the following three options:

- You receive \$5 and your counterpart receives \$5.
- Both you and your counterpart are paid based on his/her decision between "U" (\$15 for you and \$15 for him/her) and "D" (\$2 for you and \$28 for him/her).
- Your counterpart is paid according to his/her decision between "U" and "D", and you are paid based on a computer's choice between either "U" or "D".

You will not be told what the computer's decision was, or what your counterpart's decision was, unless you choose that earnings option.

---

<sup>14</sup> DONTKNOW treatment did not include the second payment choice option, KNOW treatment did not include third payment choice option and did not include the "computer's decision" paragraph (next page.)

**Room B Decision:** (The instructions given to your counterpart)

You will be anonymously assigned to a Room A counterpart who drew your number randomly from a box with the numbers 1 through 10 inside. This person will be your counterpart for the entire experiment. Your counterpart will make a decision that can affect your earnings in today's experiment. He or she can choose for both of you to be paid \$5. Another possibility is that he/she will let you determine both of your payoffs. If he/she chooses this option and you choose "U", then you get paid \$15 and he/she gets paid \$15. If you choose "D", then you get paid \$28 and he/she gets paid \$2. Your payoff will be determined in one of these two ways. Your counterpart can choose only one of the earnings methods. We will ask you to make your decision on "U" or "D" at the same time that your counterpart makes his or her choice. Your decision will only determine your payoff if your counterpart did not choose the option to give you \$5.

**Computer's Decision:**

After the Room B participants make their decisions, the computer will assign either "U" or "D" to each of the ten numbers. The computer has been programmed to assign dollar values to each of the 10 numbers in the box according to the decisions made by the Room B participants. What this means is that the number of "U" choices made by the computer is exactly the same as the number of "U" choices made by the participants in room B. Also, the number of "D" choices made by the computer is exactly the same as the number of "D" choices made by the room B participants. (Note: while the number of "U" numbers and number of "D" numbers are the same as in the Room B decisions, which numbers are assigned "U" or "D" is randomly decided by the computer) For example: if five Room B participants choose "U", then five of the numbers between 1 and 10 are randomly assigned to have the "U" payoff, and the remaining five numbers are assigned to the "D" payoff. (Note: the numbers used here are only an example and not necessarily representative of Room B decisions)



### **Room B (Trustees) Instructions: Every Treatment**

Thank you for participating in today's experiment. You've earned a \$7 show-up bonus for participating. In reading and following the instructions below, you have the potential to earn significantly more. You have been randomly assigned to **Room B**. You will also be randomly and anonymously assigned to a person in **Room A**. Your counterpart will not be told your name, and you will not be told his/her name.

You will be anonymously assigned to a Room A counterpart who drew your number randomly from a box with the numbers 1 through 10 inside. This person will be your counterpart for the entire experiment. Your counterpart will make a decision that can affect your earnings in today's experiment. He or she can choose for both of you to be paid \$5. Another possibility is that he/she will let you determine both of your payoffs. If he/she chooses this option and you choose "**U**", then you get paid \$15 and he/she gets paid \$15. If you choose "**D**", then you get paid \$28 and he/she gets paid \$2. Your payoff will be determined in one of these two ways. Your counterpart can choose only one of the earnings methods. We will ask you to make your decision on "**U**" or "**D**" at the same time that your counterpart makes his or her choice. Your decision will only determine your payoff if your counterpart did not choose the option to give you \$5.