

# Restoring Damaged Trust with Promises, Atonement and Apology

Eric Schniter <sup>a</sup>  
Roman M. Sheremeta <sup>b</sup>  
Daniel Sznycer <sup>c</sup>

<sup>a</sup>Economic Science Institute, Chapman University  
One University Drive, Orange, CA 92866

<sup>b</sup>Argyros School of Business and Economics, Chapman University,  
One University Drive, Orange, CA 92866

<sup>c</sup>Center for Evolutionary Psychology, University of California,  
Santa Barbara, CA 93106

December 22, 2011

## Abstract

In an experiment using two consecutive trust games, we study how “cheap” signals such as promises and messages are used to restore damaged trust and encourage new trust where it did not previously exist. In these games, trustees made non-binding promises of investment-contingent returns, then investors decided whether to invest, and finally trustees decided how much to return. After an unexpected second game was announced, but before it commenced, trustees could send a one-way message. This naturalistic quasi-experimental design allowed us to observe the endogenous emergence of trust-relevant behaviors and focus on naturally occurring remedial strategies used by promise-breakers and distrusted trustees, their effects on investors, and subsequent outcomes. In the first game 16.6% of trustees were distrusted and 18.8% of trusted trustees broke promises. Trustees distrusted in the first game used promises closer to equal splits and messaging to encourage trust in the second game. To restore damaged trust, promise-breakers used larger new promises (signals of intended atonement) and messaging (usually with apology). On average, investments in each game paid off for investors and trustees, suggesting that cheap signals foster profitable trust-based exchanges in these economic games.

*Keywords:* promise, atonement, apology, cheap talk, cheap signals, remedial strategies, trust game, reciprocity, experiments

---

Corresponding author: Eric Schniter, [eschniter@gmail.com](mailto:eschniter@gmail.com)

\* For inspiration to pursue this study we thank John Dickhaut. Helpful comments were received from participants at the Workshop on Communication in Games (at the University of Zurich) and the Human Behavior and Evolution Society annual meeting (in Montpellier, France). Any remaining errors are ours.

## 1.1. Introduction

In modern economies where trust realizes vast amounts of potential gains in transactions involving deferred or risky returns, problems associated with developing and restoring trust are particularly relevant. A scientific understanding of the processes that restore trust when it is damaged and encourage trust where it did not previously exist is therefore of paramount importance. Despite the large literature on damages to corporate reputation (e.g., see Barnett 2003 on US chemical industry disasters; see Robinson & Rousseau 1994 for a survey of corporate trust violations), very little research exists on how damaged trust can be rebuilt and new trust encouraged where it did not previously exist (Dirks et al. 2009). Most of the existing research in this area is either purely theoretical (Lewicki & Bunker 1996; Mishra 1996; Lewicki & Wiethoff 2000; Ren & Gray 2009; Gillespie & Dietz 2009), based on anecdotal or event-based evidence (Elsbach 1994; Knight & Pretty 1999), surveys (Slovic 1993), diary studies (Conway & Briner 2002) fabricated vignettes (Tomlinson et al. 2004), fabricated videotaped dramatizations (Kim et al. 2004, 2006), or incentivized experimental designs using deception (Gibson et al. 1999; Bottom et al. 2002; Nakyachi & Watabe 2005; Schweitzer et al. 2006; Ohtsubo & Watanabe 2009). To study how damaged trust can be rebuilt and new trust can be encouraged, we conducted a non-deceptive economic experiment with endogenously created and naturally distributed signals, using financially motivated subjects.

Our experiment is based on a version of the “investment game” by Berg, Dickhaut & McCabe (1995). In the original investment game an investor is endowed with \$10 and can invest any portion of her endowment by sending it to a trustee. The amount sent triples in value before reaching the trustee. Having received funds from this tripled investment, the trustee can reciprocate by returning any portion of these funds to the investor. Since sending money is risky, investments are usually interpreted as trust, and since returning money is costly, reciprocation via returns on investments is interpreted as evidence of trustworthiness.<sup>1</sup> The investment game, therefore, has been extensively used to study trust and reciprocity in an investment setting (for a

---

<sup>1</sup> This interpretation is based on the assumption that participants identify psychological and implied contracts (Rousseau 1989) and in doing so act in accordance with social contracts, though there is no social contract about expected or contingent behavior stated in the standard implementation of the classic “investment game” (see Berg et al. 1995), which over the years has become better known as the “trust game”. In fact, because the assertion that the original game was about “trust” was debatable, John Dickhaut preferred calling it the “investment game” – as it is in the 1995 Berg et al. article. By adding a new starting stage to the game where trustees make promises to return a portion of income from investment – this game becomes a game more explicitly about trust. For this reason we refer to our modified form of the classic investment game, described below, as a “trust game”.

review see Ostrom & Walker 2005). A common finding in the literature is that investors tend to exhibit trust and trustees tend to reciprocate. It has also been well established that pre-play communication, even if “irrelevant” to game strategy, can induce higher contributions in public goods games (for meta-analyses see Sally 1995, Balliet 2010) and more cooperation in dyadic social dilemmas (Deutsch 1958, 1960; Radlow & Weidner 1966; Buchan et al. 2002; Duffy & Feltovich 2006; Bracht & Feltovich 2009). However, with the exception of a few studies using deception, the experimental economic literature is silent as to what behavior ensues when promises fail to establish trust and what happens to trust and reciprocity in subsequent interactions after promises are broken and trust is damaged.

## 1.2. Background

Non-binding social contracts based on mutual agreement and advantage can secure opportunities to gain from trade, but may also pose risks to those entering into them: they provide cheaters opportunities for greater immediate gains while consequences to cheaters may be non-existent, uncertain, or delayed. Our research focuses on social contracts in trust-based investment exchanges that provide opportunity for mutual advantage. In these exchanges, we consider trust to be demonstrated when resources or control is willingly ceded to another with the expectation that the other intends to reciprocate. Trustworthiness is demonstrated by reciprocating so as to, at minimum, retribute the loss of resources or control that another has ceded by extending trust.

To successfully navigate a social contract and avoid exploitation by cheaters, it is important for potentially trusting investors to obtain accurate information about the ability and willingness (propensity) of the trustees to carry out their end of the contract. Trusting and trustworthy reputations that have been demonstrated by past actions serve as reliable *cues* upon which trust-based decisions can be made. Where reputational assurances are not available, such as in novel relationships with unknown partners, credible information about an investor’s willingness to trust or a trustee’s trustworthiness is not as readily accessible. In the absence of reputational cues, *signals*<sup>2</sup> are often sent to receivers with the intention to communicate

---

<sup>2</sup> We distinguish *cues* from *signals* (borrowing from similar definitions by Diggle et al. 2007; Scott-Phillips 2008) as follows. *Cue*: Any act or structure that (i) affects the behavior of other organisms; (ii) which is effective because the effect has evolved to be affected by the act or structure; but which (iii) did not evolve. *Signal*: Any act or structure that (i) affects the behavior of other organisms; (ii) evolved because of those effects; and (iii) which is effective

information about the sender (e.g., see Farrel & Rabin 1996). For example, signals may be sent with the intention of persuading receivers that the sender is more trustworthy than might be inferred from cues alone.

Signals persuading investors of a trustee's trustworthiness are fundamental to developing mutually beneficial relationships under conditions where trust has not yet been established (where cues are not available) and where trust has been damaged but not yet restored (where cues indicate untrustworthiness). Without the effective use of signals, a cooperative interaction may be foregone: potential investors may decide not to extend trust when they lack reputational assurances and when cues indicate a breach of trust. This is true whether trust has been damaged intentionally, accidentally, or as a result of mistaken interpretations of intent (Axelrod & Dion 1988).

When interests conflict (e.g., in social dilemmas), there should be skepticism about the credibility of signals of trustworthiness. Individuals may use signals to convey that they have a trustworthy propensity, yet those signals may be deceptive. When such deception brings rewards, then signalers have incentive to produce lies. If lies are a common problem, then an explanation of why recipients of those signals would continue to respond to the signals poses a problem for the evolution of signals. Natural selection would not reward recipients of dishonest signals if receiving the signals is more costly than not. Those who do not receive the same signals used to deceive others would gain a relative advantage – out-competing the unfortunate receivers of deceptive signals (Dawkins & Krebs 1978; Maynard Smith 1982). Thus, the existence of signals that can deceive selects for skepticism among potential receivers.

Zahavi (1975) partially addressed the fundamental question of “why are signals reliable?” when he suggested that some signals are reliable to the extent they are guaranteed by their costs. More specifically, the reliability of “costly” signals corresponds (positively) to the costs of their production, especially when the presence of individuals' relevant qualities is a necessary condition for offsetting the “handicapping” costs of developing or sending these signals. Insofar as signal production costs outweigh the benefits gained from using those signals deceptively (but not from using them honestly), reception of signals will continue, and deceptive senders of signals will be out-competed by honest signalers who can afford to signal (Zahavi 1977, 1993; Enquist 1985; Grafen 1990; Adams & Mesterton-Gibbons 1995). While evolutionary biologists

---

because the effect (the response) has evolved to be affected by the act or structure.

have been documenting examples of costly signaling among non-human animals,<sup>3</sup> economists have been documenting qualifying human examples from field studies and laboratory experiments.<sup>4</sup> Yet, despite the growing catalogue of “costly signal” examples, the theory used to justify their evolution only helps explain a small fraction of human signal phenomena.

We are interested in the use of less understood “cheap” signals (e.g., personalized messages, promises of reciprocation, and apologies) that do not directly affect payoffs of the game, or require substantial costs for production, yet are common features of trust-based social contracts. Personalized communication that reveals something about the sender may facilitate social contracts (Buchan et al. 2002; Ridings et al. 2002, Zheng et al. 2002) by decreasing social distance, raising solidarity, and signaling the cues of familiarity that are normally associated with trustworthy relationships. Bohnet & Frey (1999) demonstrate that personal identification, even when only one-way, leads to efficient outcomes in dyadic interactions. Promises and non-binding messages by trustees have been shown to increase cooperation (Orbell et al. 1988, Rubin & Brown 1975; Kerr & Kaufman-Gilliland 1994, Elingsen & Johannesson 2004; Charness & Dufwenberg 2006). Explanations and apologies have also been shown to have great effect on eliciting forgiveness (Lewicki & Bunker 1996; Girard & Mullet 1997; Girard et al. 2002; McCullough et al. 1997, 1998; Ohbuchi et al. 1989; Tavuchis 1991; Witvliet et al. 2002; Benoit & Drew 1997) and ensuring future trust (De Cremer et al. 2010), especially when expressing an offender’s guilt over past actions (Wubben et al. 2009) and when combined with offers to engage in atonement (Gibson et al. 1999). These remedial strategies are based on cheap signals (which presents us with the credibility problem identified by signaling theory), raising the questions of

---

<sup>3</sup> e.g. the roars of red deer, *Cervus elaphus* (Clutton-Brock & Albon 1979); spotting by Thompson’s gazelles, *Gazella thomsonii* (Fitzgibbon & Fanshawe 1988); musth in male African elephants *Loxodonta africana* (Poole 1987, 1989); and tail display of peacocks, *Pavo cristatus* (Petrie et al. 1991; Zahavi & Zahavi 1997).

<sup>4</sup> Spence (1973, 1974) has argued that the years one has spent getting an education and earning degrees signal not only intelligence, but also commitment to long-term investments and the ability to work effectively within a structured institution to prospective employers. On average, these are qualities without which degrees become too costly for unintelligent, uncommitted, or undisciplined people to get. If degrees were easy to get, they would not be reliable proxies for a worker’s propensity to be productive. Camerer (1988) suggests that an engagement ring also functions as a costly signal of a suitor’s intentions to engage in a lifetime of familial production, whereas “the lusty bachelor whose planning extends only to dawn cannot afford such costly investments” (p. S183). Gambetta (2009) describes how in prisons inmates use costly signals of “toughness” that only those with certain qualities can afford: scars from knife stabs or bullet wounds (indicating that one has been through fights yet survived), willingness to engage in fighting, and even self-inflicted harm (demonstrating one’s ability to tolerate pain). In a recently laboratory study, Fehrler and Przepiorka (2011) demonstrate that donors to charity are both expected to be and found to be more trustworthy in social exchange than non-donors.

how do people attempt using cheap signals, when do they actually "work", and who benefits from their use?

We suggest that by imposing cost (after discovering signals were false) on dishonest signalers (e.g., by excluding them from future trust-based exchanges or by spreading negative reputational information that will cause others to exclude them), receivers can make the propensity to engage in false signaling effectively “costly” enough that dishonest signalers do not gain net benefits from sending false signals. We are careful to point out that while sanctioning of false signals can reduce the frequency of their use in a population, it is not expected to drive them entirely to extinction. In fact, where opportunity costs of forgone trust-based exchange are larger, the tolerable proportion of dishonest signals to honest signals is larger. Specifically, the logic of error management theory (for a review see Haselton & Nettle 2006) predicts that despite the existence of false signaling and the costs of receiving false signals, signals will tend to be received when opportunity costs associated with not receiving true signals of trustworthiness (from forgone advantageous exchange) are greater than costs associated with receiving false signals of trustworthiness (from pursued trust-based exchange that produced a loss). The economically justified tolerance of some false signals also predicts that individuals will exploit opportunities to profit by using false signals to conceal untrustworthiness.

In sum, we argue that cheap signals can evolve based on the calculus of their production and reception costs, but that for this to happen, they should be more profitable on average to both sender and receiver than in their absence. Following this logic, we expect that cheap signals can be used to encourage new trust and restore trust that has been damaged, but that in order to do so reliably, these signals must yield relatively greater benefits to both signaler and target on average. In the laboratory, our experimental design allows us to hone in on participants’ use of the cheap signaling opportunities provided. Investigating whether these cheap signals, so important to our everyday trust-based interactions, are alive and well in the laboratory, we make several predictions. First, we predict that trustees whose actions have already produced cues establishing their trustworthy reputations (by keeping promises and not succumbing to more profitable opportunism) will be less incentivized (than previously untrusted trustees, or trustees whose reputations indicate untrustworthiness) to spend time and effort constructing messages to persuade investors to trust them. Previously untrusted trustees who have no established

trustworthiness to rely on and untrustworthy trustees (i.e., promise-breakers) are expected to make use of promises and messages to affect investors' decisions to trust. We expect that when used and "working" to affect investors' trust, signals conveying a trustworthy propensity will provide benefits to both investor and trustee on average.

### **1.3. Present Study**

Our experiment is based on a version of the "investment game" by Berg, Dickhaut & McCabe (1995). In our experiment we use two consecutive trust games to study how "cheap" signals such as promises and messages are used to restore damaged trust and encourage new trust where it did not previously exist. In these games, trustees made non-binding promises, then investors decided whether to invest, and finally trustees decided how much income to return. After the unexpected second game was announced, but before it commenced, trustees could send a one-way message. This naturalistic quasi-experimental design allowed us to observe the endogenous emergence of trust-relevant behaviors and focus on naturally occurring "cheap" remedial strategies used by promise-breakers and distrusted trustees, their effects on investors, and subsequent behaviors. In the first game 18.8% of trusted trustees broke promises and 16.6% of trustees were distrusted. Promise-breakers used cheap signals in the form of promises of larger than previously promised returns (a signal of intended atonement) and messaging (usually with apology) to restore damaged trust. Trustees who were distrusted in the first game used new promises closer to a 50/50 split and messaging to encourage trust in the second game. On average, investments paid off (for investors and trustees) in each game, netting greater earnings than non-investments.

Theories of reciprocity predict that individuals will regulate their willingness to deliver benefits (i.e., to trust) based on their expectations of another's trustworthiness. We expect that selection pressures derived from this incentive structure have put a premium on signals that convey trustworthiness, conceal untrustworthiness, and restore trustworthiness following damage to others' trust. We test our predictions from signaling theory concerning the conditions under which we expect to see cheap signals "working" to affect trust, and providing benefits for signal senders and receivers. Special attention is given to how promises are used to encourage trust where it did not previously exist and how apology and atonement can restore damaged trust.

## 2. Experimental Design and Procedures

The experiment was conducted at Chapman University's ESI laboratory. 458 participants (229 pairs) were recruited from a standard campus-wide subject pool for participation in an experiment that could last up to 45 minutes. Participants interacted with each other anonymously over a local computer network. The experiment, which lasted an average of 35 minutes total, proceeded as follows. Upon arrival, participants in the experiment were told that they would receive \$7 for participation, to be paid at the end of the experiment. Participants then received instructions (see Appendix A) for a single trust game (with no indication of a subsequent game to follow).

Subjects were assigned to one of two roles: "Participant A" for the investor or "Participant B" for the trustee. First, the trustee chose a dollar amount from \$0 to \$20 that he promised to send back to the investor, should the investor choose IN. Specifically, the trustee completed the following statement: "I (Participant B) promise to transfer back \$\_\_\_ of my income to you (Participant A) if you choose IN". This statement was not binding. That is, trustees were *not* obligated to transfer back the amount promised to the investor, and both trustee and investor knew this. The computer conveyed the trustee's statement to the investor and then the investor chose either IN or OUT. If the investor chose OUT, the investor received \$5 and the trustee \$0. If the investor chose IN, then the trustee received \$20 income. In such a case, after receiving \$20 (the "income"), the trustee chose a dollar amount from \$0 to \$20 to send back to the trustee.

After the first trust game had been completed, participants were given instructions (see Appendix A) indicating that a second trust game identical to the first would follow. In this second trust game, participants were paired with the same partner and played the same role as in the first game. However, prior to the second game, the trustee was given an opportunity to use a text box to send a one-way message to the investor (e.g., to apologize for the broken promise) and to make a new promise (e.g., to signal intended atonement). Trustees were instructed that "in these messages, no one is allowed to identify him or herself by name, number, gender, or appearance", but that other than these restrictions, trustees could "say anything in the message." If trustees wished not to send a message they were instructed to "simply click on the send button without having typed anything in the message box." The computer conveyed the trustee's message and promise to the appropriate investor, and then the second trust game began. We



specified that the second game was the last and final part of the experiment (i.e., there would be no subsequent games).<sup>5</sup>

There were 25 experimental sessions. Each session had between 10 and 24 participants. The average experimental earnings, including \$7 for arriving to the experiment on time and participating, were \$25, ranging from a low of \$7 to a high of \$47. No participant participated more than once, and no participant had prior experience with a similar game environment.

### 3. Results

#### 3.1. Game 1

We expect trustees to promise investors transfers of at least \$6 (minimally higher than the payoff to the investor if he chooses OUT), and closer to the focal point of \$10 – an even split, but less than \$20. Promises of \$20, if honored, would not provide financial benefit to the promise-makers and are therefore not expected. Wary that trustees' have less incentive to honor promises closer to \$20 than to the 50/50 split of \$10, we also expect that investors should be more suspicious of the veracity of larger promises and therefore be less likely to invest in higher promises. Trustees who have been trusted should tend to make good on their promises (or, at least, yield returns on investments that are profitable to investors on average). These predictions stand in stark contrast to the set of rational choice predictions that expect non-binding promises to have no effect on investors. According to rational choice theory, trustees who receive incomes should return nothing (despite what they may have promised) and, based on this, investors should always choose to not invest (regardless of the promise they received).

Figure 1 displays the aggregate distribution of investment and promise-keeping decisions in the experiment, while Figure 2 displays the distribution of promises made by trustees in Game 1. In Game 1, trustees on average promised to return \$9.20 (SD=2.38) out of \$20 and 83.4% (191/229) of investors chose IN.

We evaluate whether the use of Game 1 promises affected investor willingness to make trust-based investments, and whether investments made based on promises resulted in greater benefits (than non-investments) for both investor and trustee in Game 1. The distribution of promises in Figure 2 indicates that investors who chose IN received promises in the range of \$6-

---

<sup>5</sup> After each trust game subjects were also asked to fill out a 20 item survey in which they reported their emotional states consequent on their decisions, game interactions, and resulting outcomes. Analysis and discussion of the mediating roles of emotions are not included in this paper.

\$19 (99% of the time) and the most commonly received promise was for \$10 (more than 50% of the time), while the investors who chose OUT received lower promises on average (i.e., \$8.61 (SD=4.33) versus \$9.31 (SD=1.75); Wilcoxon-Mann Whitney test, p-value=0.01,  $n_1=191$ ,  $n_2=38$ ), and received either relatively high or relatively low promises overall. To confirm this observation, we estimated the effect with probit models as in Table 1 (specifications 1 and 2), where the dependent variable is the investment decision in Game 1 (*Invest1*) and the independent variables are the promise by trustee (*Promise1*) and the promise squared (*Promise1sqr*). In specification (1), the *Promise1* variable is insignificant, indicating that there is no linear relationship between the probability of investment and the promised amount. On the other hand, in specification (2), the *Promise1* and *Promise1sqr* variables are both significant, indicating that the probability of investment is significantly higher for the moderate promises (e.g., \$10 or 50% of the income).

Game 1 investments made based on promises resulted in greater benefits (than not investing) for both investor and trustee in Game 1 since investors who chose IN received back \$8.19 on average, which is substantially higher than their original endowment of \$5 (Wilcoxon signed rank test, p-value<0.01,  $n=191$ ). Compared to the \$0 earned by untrusted trustees, trusted trustees earned an average of \$11.81 in Game 1 (Wilcoxon-Mann Whitney test, p-value<0.001,  $n_1=38$ ,  $n_2=191$ ). The OLS estimation of specifications (3) and (4) in Table 1 indicates that the amount returned by trustee (*Return1*) has a non-linear relationship with the promised amount (*Promise1*, *Promise1sqr*). Specifically, returns are significantly lower for the relatively high and relatively low promises. This estimation provides evidence consistent with our prediction that those investors who chose OUT when faced with relatively low or high promises (Figure 2), would have done so out of anticipation of lower investor payoffs that we observe from investments in those promises.

For the investors who chose IN, the average amount returned of \$8.19 was significantly lower than the average promise of \$9.31 (Wilcoxon signed rank test, p-value<0.01,  $n_1=n_2=191$ ). Despite average returns being lower than promised, we find that promises of future returns tended to be veridical; 81.2% of trusted promises (155/191) were kept (i.e., the amount returned was equal to or greater than the promise), and 18.8% (36/191) were broken (i.e., the amount returned was less than the promise). Below we will refer to “promise-keepers”, meaning those who exactly kept *or* exceeded their promised returns when invested in. “Promise-breakers” will

be used below to exclusively refer to those who returned less than they promised to return when invested in (regardless of whether the return was profitable to the investor).

### 3.2. Game 2

While cheap signals are manipulated by trustees, affect investors, and provide net benefits to both investors and trustees in Game 1, facilitating profitable trust-based exchanges where previous reputations had not been established, Game 2 provides us a relatively different game environment in which to study cheap signals. In Game 2, reputations have been established for 83.4% of trustees (in terms of demonstrated trustworthiness), and 100% of investors (in terms of demonstrated trust) – raising the question of whether the use of cheap signals will still matter where cues of willingness to trust and act trustworthy have been established.

On average, trustees promised to return \$9.79 in Game 2, a larger amount than the average of \$9.20 promised in Game 1 (Wilcoxon signed rank test,  $p$ -value $<0.01$ ,  $n=229$ ), which resulted in 87.3% (200/229) of investors choosing IN, only slightly more than the 83.4% (191/229) of IN decisions made in Game 1 (Fisher's exact test,  $p$ -value=0.59,  $n=229$ ). We consider whether investor and trustee reputations established in Game 1 and the new promises issued in Game 2 affect investment decisions. The estimation of probit models in Table 2 (specifications 1 and 2) indicates that the decision to invest in Game 2 (*Invest2*) mainly depends on the promise in Game 2 (*Promise2*, *Promise2sqr*), with no significant effects found for Game 1 distrusted versus trusted trustees (*Distrusted1*) or for Game 1 promise-keepers versus promise-breakers (*Broken1*). In the sections below we further explore the effect of promises and messages on Game 2 investments within the subsamples aggregated by Game 1 decisions.

We evaluate whether the use of Game 2 promises and the extent of promise-breaking in Game 1 affected investor willingness to make trust-based investments, and whether investments made based on new promises and the extent to which promises were previously broken resulted in greater benefits (than from non-investment) for both investor and trustee in Game 2. Overall, the investments made in Game 2 again paid off since their investors received an average \$8.73 return from their investment choice IN, which is significantly higher than the OUT payoff of \$5 (Wilcoxon signed rank test,  $p$ -value $<0.01$ ,  $n=200$ ). The estimation of specifications (3) and (4) in Table 2 indicates that, similar to Game 1, returns in Game 2 (*Return2*) depend on promises made in Game 2, although linearly this time (*Promise2*). In addition, returns negatively depend on the

extent of the broken promise in Game 1 (*Promise1-Return1*), suggesting that trustees' extent of untrustworthiness (defined by the amount which a return was less than promised) in Game 1 is predictive of earnings that investors can expect in Game 2. Overall, similar to Game 1, promises of profitable returns on investment in Game 2 tended to be veridical; 75% of promises (150/200) were kept or exceeded, and 25.0% (50/200) were broken. In the sections below we further explore the effect of promises and messages on Game 2 earnings within the subsamples aggregated by Game 1 decisions.

### 3.2.1. Game 1 Promise-Keepers

For the subset of 155 promise-keeping trustees (i.e., those who did not break their promises in Game 1), we observe slightly higher average promises in Game 2. Figure 3 displays the histogram of promises made in Game 2 by 155 promise-keepers from Game 1. These trustees promised to return an average of \$9.46 in Game 2, which is higher than their average promise of \$9.02 in Game 1 (Wilcoxon signed rank test,  $p\text{-value} < 0.01$ ,  $n_1 = n_2 = 155$ ).

Perhaps as a consequence of Game 2 promises close to 50/50 splits of income, 92.3% of Game 2 investors in Game 1 promise-keepers (143/155) chose IN. Note that the investment rate of 92.3% is higher (Wilcoxon-Mann Whitney test,  $p\text{-value} = 0.007$ ,  $n_1 = 155$ ,  $n_2 = 229$ ) than the investment rate of 83.4% in Game 1 (191/229). While this rate increase in trust may be explained in part by updated promises in Game 2 (\$9.46 versus \$9.02), it can also be explained by the profitable returns transferred, which for promise-keepers was always the amount that they promised or more. Specifications (1) and (2) in Table 3 support the conjectures that new Game 2 promises and Game 1 returns by promise-keepers affect Game 2 investments.<sup>6</sup> Specifically, specification (1) indicates that Game 2 investment decisions (*Invest2*) are positively correlated with returns in Game 1 (*Return1*) and non-empty messages (*Message*). Specification (2) indicates that when promises are updated nonlinearly, Game 2 investment decisions (*Invest2*) are positively correlated with updated Game 2 promises (*Promise2*, *Promise2sqr*) and non-empty messages (*Message*).

We predicted that – due to their established reputations of trustworthiness (as compared to either promise-breakers or distrusted trustees) – Game 1 promise-keepers were relatively less

---

<sup>6</sup> Note that in estimating these regressions we cannot include both *Return1* and *Promise1* since for promise-keepers they are perfectly collinear.

incentivized to spend time and effort constructing messages to persuade investors to choose IN in Game 2, and so would send both shorter messages and a greater frequency of empty messages in Game 2. Game 1 promise-keepers' messages contained fewer words than messages from both Game 1 untrusted trustees and Game 1 promise-breakers ( $M=11.41$  versus  $M=22.9$ , Wilcoxon rank-sum test,  $p<0.01$ ,  $n_1=155$ ,  $n_2=74$ ). Comparatively, Game 1 promise-keepers' messages were also more frequently empty (20% versus 11% of the time, Wilcoxon rank-sum test,  $p=0.06$ ,  $n_1=155$ ,  $n_2=74$ ).

We evaluate whether investments made based on new promises and non-empty messages resulted in benefits for both investor and trustee in Game 2, and whether these signals were reliable indicators of subsequent trustee behaviors. Investments in Game 1 promise-keepers paid off for investors choosing IN in Game 2. These investors received an average of \$8.62 from trustees, as opposed to the \$5 earned from OUT (Wilcoxon signed rank test,  $p\text{-value}<0.01$ ,  $n=143$ ), with 83.9% (120/143) of the promises kept or exceeded, and 16.1% (23/143) broken. Compared to \$0 earned by untrusted Game 1 promise-keepers in Game 2, promise-keepers also profited from trusted promises in Game 2 earning \$11.38 on average (Wilcoxon-Mann Whitney test,  $p\text{-value}<0.01$ ,  $n_1=12$ ,  $n_2=143$ ). Specification (3) in Table 3 indicates a positive linear relationship for Game 2 returns (*Return2*) and promises (*Promise2*), again further confirming that the trusted promises are reliable cues of returns (which are profitable on average). No significant effect of promise-keepers messages (*Message*) was seen on Game 2 returns (*Return2*).

### 3.2.2. Game 1 Promise-Breakers

A major question our data address concerns what happens when a fresh opportunity for cooperation arises between two parties subsequent to a violation of trust: how trustees react, how investors respond, and what outcomes are achieved. Here we focus on the 18.8% (36/191) of pairs where promises were broken in Game 1 (i.e., where the amount returned was lower than the promise). These broken promises represent breaches of trust, and the relationships that immediately follow are considered to have *damaged trust* (i.e., because trust-based expectations were not met). A central question motivating this study is, “can cheap signals restore the trust damaged by broken promises in a previous interaction?”

We consider larger new promises (i.e., where  $Promise2 > Promise1$ ) made by Game 1 promise-breakers to be intentional remedial strategies which we call *promises of intended atonement*. If honored, returns from a promise of intended atonement can be construed as contributions towards restitution of the previously promised amount, expected but lost when *Promise1* was trusted and broken. Figure 4 displays the histogram of promises made in Game 2 by 36 promise-breakers. Promise-breakers promised \$12.11 in Game 2, which is significantly higher than their promise of \$10.58 in Game 1 (Wilcoxon signed rank test, p-value=0.01,  $n_1=n_2=36$ ) – signaling intended atonement (i.e.,  $Promise2 - Promise1 > 0$ ). It appears that a signal of intended atonement partially restores trust, since 69.4% (25/36) of investors whose trust was damaged in Game 1 chose IN again. While promises of intended atonement could signal intention to provide an “economic” contribution towards restituting the previously promised amount lost, victims of damaged trust might also require additional indications that the trustee has changed his investor-regarding disposition before re-extending trust and again choosing IN.

In addition to larger new promises, we also find that Game 1 promise-breakers frequently used messages whose features (see below) we assume were intentionally manipulated to persuade investors to choose IN in Game 2. Table B1 in Appendix B reports all messages that were sent by 36 promise-breakers. Analyzing the messages, we find that 83.3% (30/36) of the messages have some content. Game 1 promise-breakers’ messages contain more words than messages from Game 1 promise-keepers (M=19.1 versus M=11.4, Wilcoxon rank-sum test, p=0.03,  $n_1=36$ ,  $n_2=155$ ) whom we assume – due to their established reputations of trustworthiness – were not as incentivized to construct a message with content for the purpose of persuading investors to choose IN in Game 2. Furthermore, we find that 80% (24/30) of messages with content restore trust (i.e., investors choose IN in Game 2 after suffering broken promises in Game 1), as opposed to only 16.7% (1/6) of messages without content. These differences are significant (Fisher’s exact test, p-value<0.01, n=36). These observed differences in larger new promise and message style indicate that most Game 1 promise-breakers made use of both a signal of intended atonement (a larger new promise) and personalized communication (a one-way ad libitum message) to restore damaged trust and persuade investors to re-trust them.

We expect that the use of apologies (i.e., with remorse, regret, or sorrow stemming from acknowledgment of offense)<sup>7</sup> should increase investors' willingness to reinvest in promise-breakers. Out of 30 messages with content, we coded 10 messages as apologies.<sup>8</sup> We find that 90.0% (9/10) of apologizers were retrusted in comparison to only 61.5% (16/26) of non-apologizers (Fisher's exact test, p-value=0.10, n=36), indicating that messages with apology are more likely to restore trust after broken promises than empty messages or messages without apology.

To study the link between messages expressing regret for an offensive action (apology) and intentions to demonstrate atonement (new larger promises) we evaluate whether apologies issued in the experiment actually correlated with larger promises of intended reciprocations. Among Game 1 promise-breakers, the restitution promised (i.e., *Promise2-Promise1*) by signals of intended atonement is significantly higher for those participants who issued an apology than for those who did not (\$3.00 versus \$0.65; Wilcoxon rank-sum test, p-value=0.06, n<sub>1</sub>=10, n<sub>2</sub>=20). Promise-breakers who sent a non-apology message (20/30) increased promises by \$0.65 (SD=3.20), while promise-breakers who sent an apology message (10/30) increased promises by \$3.00 (SD=2.62). When compared to the whole population of trustees, the difference is even more striking. For 10 trustees who issued apologies the increase in promises is more than six times higher than for all other 219 trustees (\$3.00 versus \$0.48; Wilcoxon rank-sum test, p-value<0.01, n<sub>1</sub>=10, n<sub>2</sub>=219), indicating that apologetic trustees increased their second promises more than all other trustees.

---

<sup>7</sup> There are two commonly accepted definitions of *apology*, one broader and one more narrow. Throughout this paper, we focus on the narrower definition unless we specify otherwise. The broader definition of the word *apology* comes from the Late Latin *apologia*, "a speech in defense", which itself derives from Greek *apologos* "an account, story" (*apo*: "from, off"; *logos*: "speech"). According to Webster's Revised Unabridged Dictionary (Porter 1913) "An *apology*, in the original sense of the word, was a *pleading off* from some charge or imputation, by explaining and defending one's principles or conduct. It therefore amounted to a vindication. One who offers an *apology*, admits himself to have been, at least apparently, in the wrong, but brings forward some palliating circumstance, or tenders a frank acknowledgment, by way of reparation." The more recent Merriam Webster Online Dictionary (2011) defines apology first more broadly as "1a: a formal justification, b: excuse", and second, more narrowly, as "an admission of error or discourtesy accompanied by an expression of regret". Other sources indicate that the more commonly accepted definition is the narrow one. For example, the American Heritage Dictionary (2011) first defines apology as "Written or spoken expression of one's regret, remorse, or sorrow for having insulted, failed, injured, or wronged another". Conversely, American Heritage's second definition, "a defense, excuse, or justification in speech or writing, as for a cause or doctrine", is consistent with Webster's first.

<sup>8</sup> 23 messages can be qualified as apologies in the broader sense (i.e., an explicit or implicit acknowledgment of another's offense received which, in the context of this experiment, is a trusted promise that was broken). Using broader definition of apologies, we find no substantial differences in investment rates between apology and no apology (78.2% versus 85.7%).

Thus far, we have only considered the independent effects of intended atonement (new larger promises) and messages in restoring damaged trust, but recognize that these remedial strategies are often used together. Next, we estimate probit regressions as in Table 4 to identify how these remedial strategies work in conjunction. Specification (1) indicates that the two most significant predictors of trust in Game 2 (*Invest2*) are new larger promises (*Promise2-Promise1*) and non-empty messages (*Message*). Specification (2) shows that in addition trust is negatively affected by the magnitude of broken promise in Game 1 (*Promise1-Return1*). These results indicate that investors respond to the combined effects of adjusted promises and longer messages by making trust-based investments in Game 2 in previously distrusted trustees.

Evolutionary theory argues that signals like apologies and promises of intended atonement should have evolved only if they provided net benefits to both the senders and receivers of the signals. We evaluate whether Game 1 promise-breakers' signals resulted in benefits for both investor and trustee in Game 2, and whether these signals were reliable indicators of subsequent trustee behaviors. Investors in Game 1 promise-breakers were returned on average \$7.28, which is significantly higher than the OUT payoff of \$5 (Wilcoxon signed rank test,  $p$ -value=0.05,  $n=25$ ). Moreover, Game 1 promise-breakers returned significantly more in Game 2 than in Game 1 (\$7.28 versus \$4.60; Wilcoxon signed rank test,  $p$ -value<0.01,  $n_1=n_2=25$ ). This is also true when we look at investments in the subset of 9 out of 10 trustees who explicitly issued apologies and where retrusted (\$6.78 versus \$4.22; Wilcoxon signed rank test,  $p$ -value=0.06,  $n_1=n_2=9$ ). Although on average investments in Game 2 paid off, we still find that 60.0% (15/25) of trustees who broke their promises in Game 1 (and were subsequently retrusted), broke their promises again in Game 2 – almost irrespective of the apologies and new promises. Apologies were not veridical on average;<sup>9</sup> only 4/9 (44.4%) retrusted apologizers kept Game 2 promises,<sup>10</sup> a greater but not significantly different (Fisher's exact test,  $p$ -value=0.53,  $n=25$ ) proportion than 6/16 (37.5%) retrusted non-apologizers who kept Game 2 promises. From specifications (3) and (4) in Table 4, it appears that the most significant predictor for return in Game 2 (*Return2*) from a promise breaker is a sent message with content (*Message*).

### 3.2.3. Game 1 Distrusted

---

<sup>9</sup> Likewise, of the 18 more broadly defined apologies that were retrusted, only 44.4% were veridical.

<sup>10</sup> 8 of these 9 apologizers also signaled intended atonement.



As mentioned, 38 out of 229 trustees (16.6%) were not trusted in Game 1 (see Figure 1). We attribute this distrust to the fact that these trustees offered either relatively high or relatively low promises (see right panel of Figure 2). In particular, in Game 1, 55.3% (21/38) of distrusted trustees promised less than \$9 while another 10.5% (4/38) of them promised more than \$11. As with our Game 1 predictions of trusted promises, we expect that distrusted trustees would adjust their promises towards the modal and more trusted promise of \$10, that these adjustments would affect decisions to invest, and that investments made based on adjusted promises would benefit both the investor and trusted.

First we evaluate whether Game 1 distrusted trustees adjusted their promises as we expected and if adjustments of promises by Game 1 distrusted trustees affect investment decisions. Trustees who were distrusted in Game 1 promised an average of \$8.92 in Game 2, which is similar to their average promise of \$8.61 in Game 1 (Wilcoxon signed rank test,  $p$ -value=0.45,  $n_1=n_2=38$ ), yet most investors (84.2% or 32/38) who did not trust in Game 1 chose IN in Game 2. Figure 5 displays the histogram of promises made in Game 2 by the 38 trustees who were distrusted in Game 1. Distrusted trustees changed their distribution of promises towards more equal splits: 66.7% (14/21) of trustees who promised less than \$9 in Game 1 increased their Game 2 promises and 100% (4/4) of trustees who promised more than \$11 in Game 1 decreased their Game 2 promises. Correspondingly, among previously un-trusting investors, 92.6% (13/14) of those who received increased promises and 100% (4/4) of those who received the decreased promises chose IN in Game 2.

Next, we analyze whether new trust in previously distrusted trustees can be statistically attributed to how distrusted trustees utilized messages and recalibrated promises. We expected that distrusted trustees would construct longer messages with content (and be more incentivized to do so than trustees who had already established reputations of trustworthiness) to persuade investors to choose IN in Game 2. Table B2 in Appendix B reports the messages that were sent by 38 trustees who were distrusted in Game 1. Analyzing these messages, we find that 94.7% (36/38) of the messages used by distrusted trustees have some content. Game 1 distrusted messages contain more words than messages from Game 1 promise-keepers ( $M=26.6$  versus  $M=11.4$ , Wilcoxon rank-sum test,  $p<0.01$ ,  $n_1=38$ ,  $n_2=155$ ). These data suggest that distrusted trustees use both promises adjusted towards 50/50 divisions of income and longer messages to persuade investors to trust them. The estimation of specification (1) in Table 5 indicates that the

investment decisions in Game 2 (*Invest2*) are positively correlated with new up-regulated promises in Game 2 (*Promise2sqr*) and the length of the message (*Wordcount*), indicating that investors respond to the adjusted promises and longer messages used by distrusted trustees by making trust-based investments in Game 2.<sup>11</sup> Overall, in Game 2, after trustees sent messages and updated their promises, 84.2% (32/38) of previously untrusting investors chose IN. This is very similar to the original investment rate of 83.4% in Game 1.

Finally, we evaluate whether the cheap signals successfully used by Game 1 distrusted trustees' to build new trust resulted in benefits for both investor and trustee in Game 2, and whether these signals were reliable indicators of subsequent trustee behaviors. Game 2 investments made in previously distrusted trustees paid off for investors and trustees. Investors in Game 1 distrusted trustees were returned on average \$6.88, which is significantly higher than the OUT payoff of \$5 (Wilcoxon signed rank test, p-value=0.05, n=32). Trustees who were distrusted in Game 1, but then trusted in Game 2, kept or exceeded their promises 62.5% of time (20/32). Nevertheless, 37.5% (12/32) of previously distrusted trustees who were trusted in Game 2 broke their promises – more (Wilcoxon rank-sum test, p=0.048, n<sub>1</sub>=32, n<sub>2</sub>=191) than the 18.8% of trusted trustees who broke their promises in Game 1. The estimation of specification (2) in Table 5 shows very weak correlation between Game 1 distrusted trustees' returns in Game 2 (*Return2*) and their promises (*Promise2*, *Promise2sqr*), suggesting that the extent to which the 37.5% of Game 1 distrusted trustees break their Game 2 promises is noteworthy. It is possible that the some newly trusted trustees who broke their promises in Game 2 did so in order to punish investors for their distrust in Game 1. By doing so these presumed punishers ended up earning an average of \$17.42 in two games, closer to the average earning of \$21.99 across two games for Game1 trusted trustees, than the average earnings of \$10.55 for newly trusted trustees who did not break promises in Game2.

#### 4. Discussions and Conclusions

Opportunities for mutual gains often exist where previous trust-based exchange histories have not yet been developed, or where trust has been damaged by a failure to meet expectations. Our natural experiment demonstrates that in these situations people use (i.e. send and receive)

---

<sup>11</sup> The variable *Message* is omitted from estimation of specification (1) because log-likelihood does not converge as the variable *Message* is almost perfectly collinear with *Constant*.

cheap signals to encourage new trust and rebuild damaged trust, despite the risks that these signals may be dishonest.

With promises used to encourage new trust in Game 1, and among untrusted trustees in Game 2, we see that IN decisions by investors are higher when promised returns are within the range that would provide benefits to both signal sender and receiver, and especially when the promised split is even (where conflict between individual incentives is balanced). We suggest two reasons why non-binding promises were effectively used to establish new trust in our experiment.

First, if the promises themselves are not costly, then the credence they are given by investors may be driven by a default assumption that future exclusion of cheaters and the spread of negative gossip concerning their untrustworthy reputations will make the cost of defection high enough (e.g., see Delton et al. 2011). For investors in our games, what is certain is that the investment decision may be the last investment decision (because Game 1 is describe as a single interaction with no indication that future interactions in the pairing are to be expected, and the unexpected Game 2 is described as a single interaction with assurance that no further games will occur). Investors' ability to use exclusion based on discovery of broken promises is therefore not a certainty in either interaction. Yet, despite the propositional information of these games, the mind is not designed for terminal anonymous interactions. We expect that evolved psychologies bring psychological contracts with an assumption of excludability into the lab, which lends credence to the signals (but not for reasons provided by the game environment). While we find this argument convincing, we do not present direct evidence to support it.

Second, despite promises not being costly (i.e., dishonest promise makers will profit when trusted), potential receivers of these signals need to maximize their tradeoffs between costs of type I errors (i.e., losses of endowment from making trust-based investments based on promises that are not honest) and the costs of type II errors (i.e., losses from foregone returns from not making trust-based investments based on distrusting promises that were honest). Where opportunity costs of forgone trust-based exchange are larger, the tolerable proportion of dishonest signals to honest signals is larger. Unlike the standard investment game that uses a multiplier of 3, we used a multiplier of 4 – which ultimately created a large opportunity cost for investors who chose OUT. Game 1 Investors who chose IN received back \$8.19 on average, which is substantially higher than their original endowment of \$5. Despite the rate of broken

promises observed in Game 1, trusted promises produced more profits for investors than untrusted promises. We suspect that the signaling psychology used by senders is sensitive to the tradeoffs considered by investors as evidenced by the non-random distribution of promises made and the correspondence of this distribution to promises trusted. Likewise we suspect that the extent to which promises are broken is a product of the net profits to investors, the amount originally promised, and profits to promise breakers.

In Game 2, evidence indicates that cheap signals are manipulated by trustees and affect investors, even (and especially) under conditions where cues of untrustworthiness have been established. As we expected, messaging is not utilized effectively by Game 1 promise-keepers, but is used with intended effect by those who previously broke promises as well by previously untrusted trustees. Promise breakers specifically crafted apologies to acknowledge past offenses, express remorse and regret, and persuade investors that if invested in again, things would go better for investors. As we discussed for Game 1 promises, messages are cheap signals whose reliability is not guaranteed by their direct costs. We suspect that participants brought evolved psychologies into the lab which evaluated the credence of messages “as if” the default assumption that future exclusion of cheaters and the spread of negative gossip concerning their untrustworthy reputations would make the cost of defection high enough. Again, we do not have evidence to directly support this conjecture, but we do have evidence that the decision to invest in promise breakers and previously untrusted trustees is statistically correlated with the message use, and that these investments were more profitable for the investors and trustees than the decision to not invest would have been. When the sending and receiving of cheap signals is profitable to sender and receivers, cheap signals are adaptive and expected to exist at evolutionary equilibrium. From this study we see evidence indicating how personal exchanges are often based around establishment of trust via cheap signals, and how these cheap signals can encourage new trust where it did not previously exist or repair trust where it had been damaged. Not only is this important information that could improve understanding of what to expect from our everyday interpersonal relationships, it is information that complements our understanding of how market exchange systems (where interactions often take place between non-personal entities such as firms), politics, law, and religion are sometimes expected to work, with personal representatives making verbal and written promises of reciprocation or atonement or else issuing apologies and personalized messages. Both interpersonal interactions and markets are built on

the ancient human foundations of adaptive giving and receiving. As such, trust-based exchanges at any level are often based around establishment of trust via cheap signals such as claims about reputation, verbal contracts, and apologies.

While the persuasive effects of apology and signals of intended atonement on restoring damaged trust have been clearly demonstrated in this lab experiment, it is important to note that the majority (60%) of investors whose trust was damaged in Game 1 and were persuaded to trust again in Game 2 were, again, met with broken promises. This raises the question of why humans might be so easily persuaded by cheap talk like apology and atonement. Our argument that future excludability supports assigning credence to signals predicts that if we would allow tertiary interactions among our participant pairings, repeated promise-breaking should make investors devalue cheap signals like promises and apologies and choose OUT in future rounds. We also suggest that for the 60% of trustees whose trust had been initially damaged, and then re-extended trust based on apologies only to have their trust damaged again, –may not have occurred outside of the laboratory where emotional states are reliably communicated through other forms simultaneously (e.g., facial expressions, voice, body language) and in concert with additional reputational information and opportunities for sanctioning undesirably behavior. We suspect that in the “real world” of non-anonymous and face-to-face interactions, persuasive messages like apology and promises of intended atonement are likely more reliable and less likely to lead to further damaged trust because the message receiver can evaluate the veracity of a verbal message according to not only internal coherence among cheap talk signals (a lack of which might demonstrate intentional lying), but also the correspondence of verbal signals with other reliable signals (e.g., facial expressions, past demonstrations of trust or trustworthiness, tone of voice, eyes, body language).<sup>12</sup>

Based on our findings and a review of the current literature we suggest three steps that can be taken as a remedial strategy to restore damaged trust. First, when trust in a relationship has been damaged, the offender should recognize the damage, empathize with the victim’s perspective, and communicate a desire to implement change in the relationship. An optimistic perspective on relationships fraught with damaged trust recognizes that they actually represent

---

<sup>12</sup> Hirshleifer (1984) theorized that emotions act as “guarantors of threats and promises” and several authors (Van Kleef et al. 2004, 2006; Sinaceur & Tiedens 2006, Wubben et al. 2008; Stouten & De Cremer 2010) have demonstrated experimentally that displays of emotion (including anger, guilt, happiness, disappointment, worry, regret) are used by observers for subsequent decision making in social dilemmas and negotiations.

opportunities to develop better relations than previously established. Second, to persuade and assure victims that relationship repair is possible the offender must signal seriousness, commitment, and give indication of the value that is recognized in the other, which is the basis and motivation for actual change to come. In signaling recognition of relationship value it is important not to express a selfish welfare perspective, but instead an other-regarding or shared welfare perspective. Third, to actually begin the process of changing and redefining the relationship, an offender must be willing to expeditiously take on costs by either sacrificing wealth or status, or by taking action to correct the previous imbalance of welfare that was realized by the transgression. When corrective actions cannot be taken, signals of intent to take corrective actions should be used. These three steps are identified as each having independent effects of improving impressions of the offender (Scher & Darley 1997; Schlenker 1980) and are consistent with the proscriptions detailed by De Cremer (2010) for the financial world to restore their damaged trust with customers, as well as the conclusions that Lazare (2004) arrived at through thorough analysis of how apologies are used (and misused) across applications and their relative efficacy.

As the natural occurrence of deceit in social exchanges is sampled and the effectiveness of strategies, tools, and institutions used to combat it are evaluated, practical insights are gleaned that can be extended to our personal lives, to the work of policy makers, and even applied to the handling of firms and industry affairs. We strongly encourage further efforts to uncover effective strategies for building up trust where previous trust-based exchange histories had not been developed, or where trust had been damaged by reciprocation failure.

## References

- Adams, E. S. & Mesterton-Gibbons, M. (1995). The cost of threat displays and the stability of deceptive communication. *Journal of Theoretical Biology*, 175: 405–421.
- apology. (2011a). In *Merriam-Webster Online Dictionary*. Retrieved May 6, 2011, from <http://www.merriam-webster.com/dictionary/apology>
- apology. (2011b). In *American Heritage Online Dictionary*. Retrieved May 6, 2011, from <http://www.ahdictionary.com/word/search.html?q=apology>
- Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science* 242:1385-1389
- Barnett, M. (2003). Unringing the Bell: Can Industries Reverse Unfavorable Institutional Shifts Triggered by Their Own Mistakes? *Southern Management Association Conference Proceedings* pp.800-806
- Balliet, D. (2010). Communication and cooperation in social dilemmas: a meta-analytic review. *Journal of Conflict Resolution* 54(1): 39-57
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior* 10:122-142
- Benoit, W.L., & Drew, S. (1997). Appropriateness and effectiveness of image repair strategies. *Communication Reports*, 10:153–163.
- Bohnet, I. & Frey, B.S. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior & Organization* 38:43-57.
- Bottom, W., Daniels, S., Gibson, K.S., & Murnighan, J.K. (2002). When talk is not cheap: Substantive penance and expressions of intent in the reestablishment of cooperation. *Organization Science*, 13: 497-515.
- Bracht, J. & Feltovich, N. (2009). Whatever you say, your reputation precedes you: observation and cheap talk in the trust game. *Journal of Public Economics* 93:1036-1044.
- Buchan, N., Johnson, E., & Croson, R. (2002) Trust and reciprocity: An international experiment. mimeo, Wharton School, University of Pennsylvania.
- Camerer, C. F. (1988). Gifts as economic signals and social symbols. *American Journal of Sociology* 94:180-214.
- Charness, G. & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica* 74:1579-1601.
- Clutton-Brock, T. H. & Albon, S. D. (1979). The roaring of red deer and the evolution of honest advertisement. *Behaviour*, 69:145–170.
- Conway, N., & Briner, R. B. (2002). A daily diary study of affective responses to psychological contract breach and exceeded promises. *Journal of Organizational Behavior*, 23: 287–302.
- Dawkins, R. & Krebs, J. R. (1978). Animal signals: information or manipulation. In: *Behavioural Ecology: an Evolutionary Approach*. 1st edition (Ed. by J. R. Krebs & N. B. Davies), pp. 282–309. Oxford: Blackwell Scientific.
- Deutsch, Morton. (1958). Trust and suspicion. *Journal of Conflict Resolution* 2:256-79.
- Deutsch, Morton. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations* 13:123-3
- De Cremer, D. (2010). Rebuilding trust. *Business Strategy Review*, 21(2), 79-80.
- De Cremer, D., van Dijk, E., & Pillutla, M.M. (2010). Explaining unfair offers in ultimatum games and their effects on trust: an experimental approach. *Business Ethics Quarterly* 20(1):107-126.

- Delton, A. W., Krasnow, M. M., Cosmides, L. & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 13335-13340.
- Diggie, S.P., Gardner, A., West, S.A. & Griffin, A.S. (2007). Evolutionary theory of bacterial quorum sensing: when is a signal not a signal? *Philosophical Transactions of the Royal Society B* 362:1241–1249.
- Dirks, K. T., Lewicki, R. J., & Zaheer, A. (2009). Repairing relationships within and between organizations: Building a conceptual foundation. *Academy of Management Review*, 34(1), 68–84.
- Duffy, J. & Feltovich, N. (2006). Words, deeds, and lies: strategic behavior in games with multiple signals. *Review of Economic Studies* 73:669-688.
- Elingsen, T., & Johannesson, M. (2004). Promises, threats, and fairness. *Economic Journal* 114:397-420.
- Elsbach, K. D. (1994). Managing organizational legitimacy in the California cattle industry: The construction and effectiveness of verbal accounts. *Administrative Science Quarterly*, 39: 57–88.
- Enquist, M. (1985). Communication during aggressive interactions with particular reference to variation in choice of behaviour. *Animal Behaviour*, 33:1152–1161.
- Farrell, J. & Rabin, M. (1996). Cheap Talk. *Journal of Economic Perspectives* 10 (3): 103–118.
- Fehrler, S., & Przepiorka, W. (2011). Charitable giving as a signal of trustworthiness: disentangling the signaling benefits of altruistic acts. *Mimeo* version 21/07/2011.
- Fitzgibbon, C. D. & Fanshawe, J. H. (1988). Stotting in Thompson's gazelles: an honest signal of condition. *Behavioral Ecology and Sociobiology*, 23:69–74.
- Gambetta, D. (2009). *Codes of the Underworld: How Criminals Communicate*. Princeton: Princeton University Press.
- Gibson, K., Bottom, W., & Murnighan, K. (1999) Once bitten: Defection and reconciliation in a cooperative enterprise. *Business Ethics Quarterly*, 9: 69-85.
- Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review*, 34: 127–145.
- Girard, M., & Mullet, E. (1997). Propensity to forgive in adolescents, young adults, older adults, and elderly people. *Journal of Adult Development*, 4, 209–220.
- Girard, M., Mullet, E., & Callahan, S. (2002). The mathematics of forgiveness. *American Journal of Psychology*, 115, 351–375.
- Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, 144:517–546.
- Haselton, M.G. & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality & Social Psychology Review* 10(1):47-66
- Hirshleifer, J. (1984). On emotions as guarantors of threats and promises. *UCLA Department of Economic Working Paper #337* (August).
- Kerr, N. & Kaufman-Gilliland, C. (1994). Communication, commitment, and cooperation in social dilemmas. *Journal of Personality and Social Psychology* 66:513-529.
- Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, 99: 49–65.



- Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The effects of apology vs. denial for repairing ability- vs. integrity-based trust violations. *Journal of Applied Psychology*, 89: 104–118.
- Knight, R.F. & Pretty, D.J. (1999). Corporate catastrophes, stock returns, and trading volume. *Corporate Reputation Review*, 2(4):363-378
- Lazare, A. (2004). *On Apology*. New York: Oxford University Press
- Lewicki, R.J., & Bunker, B.B. (1996). Developing and maintaining trust in working relationships. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research*. Thousand Oaks, CA: Sage.
- Lewicki, R.J. & Wiethoff, C. (2000). Trust, trust development, and trust repair. In M. Deutsch & P. T. Coleman (Eds.), *Handbook of conflict resolution: Theory and practice*. San Francisco, CA: Jossey-Bass.
- Maynard-Smith, J. (1982). Do animals convey information about their intentions? *Journal of Theoretical Biology*, 97:1–5.
- McCullough, M.E., Worthington, E.L., Jr., & Rachal, K.C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, 73, 321–336.
- McCullough, M.E., Rachal, K.C., Sandage, S.J., Worthington, E.L., Brown, S.W., & Hight, T.L. (1998). Interpersonal forgiving in close relationships. II: Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, 75, 1586–1603.
- Mishra, A. K. (1996). Organizational responses to crisis: The centrality of trust. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research*: 261– 287. Thousand Oaks, CA: Sage
- Nakayachi, K., & Watabe, M. (2005). Restoring trustworthiness after adverse events: The signaling effects of voluntary “hostage posting” on trust. *Organizational Behavior and Human Decision Processes*, 97: 1–17.
- Ohbuchi, K., Kameda, M., & Agarie, N. (1989). Apology as aggression control: Its role in mediating appraisal of and response to harm. *Journal of Personality and Social Psychology*, 56: 219-227.
- Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior*, 30(2), 114-123.
- Orbell, J., Dawes, R., & Kragt, A. van de (1988). Explaining discussion induced cooperation. *Journal of Personality and Social Psychology*, 54: 811-819.
- Ostrom, E. & Walker, J. (2005). *Trust and reciprocity: Interdisciplinary lessons from experimental research*. Russell Sage Foundation Publications.
- Petrie, M., Halliday, T., & Sanders, C. (1991). Peahens prefer peacocks with elaborate trains. *Animal Behavior* 141:323-331.
- Poole, J. H. (1987). Rutting behaviour in African elephants: the phenomenon of musth. *Behaviour*, 102:283–316.
- Poole, J. H. (1989). Announcing intent: the aggressive state of musth in African elephants. *Animal Behaviour*, 37:40–152.
- Porter, N. (editor). (1913). *Webster's Revised Unabridged Dictionary*. G & C. Merriam Co.
- Radlow, R. & Weidner, M.F. (1966). Unenforced commitments in “cooperative” and “noncooperative” non-constant-sum games. *Journal of Conflict Resolution* 10:497-50
- Ren, H., & Gray, B. (2009). Repairing relationship conflict: How violation types and culture influence the effectiveness of restoration rituals. *Academy of Management Review*, 34: 105–126

- Ridings, C.M., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information Systems* 11 (3), 271–295.
- Robinson, S. L., & Rousseau, D. M. (1994). Violating the psychological contract: Not the exception but the norm. *Journal of Organizational Behavior*, 15: 245-259.
- Rousseau, D.M. (1989). Psychological and implied contracts in organizations. *Employee Responsibilities and Rights Journal*, 2(2):121-139.
- Rubin, J. & Brown, B. (1975). *The Social Psychology of Bargaining and Negotiation*. New York, NY: Academic Press.
- Sally, D. (1995). Conversation and cooperation in social dilemmas. *Rationality and Society* 7(1), 58-92
- Schweitzer, M.E., Hershey, J., & Bradlow, E. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes*, 101: 1–19.
- Scott-Phillips, T.C. (2008). *Defining biological communication*. *Journal of Evolutionary Biology*, 21:387-395
- Scher, S.J., & Darley, J.M. (1997). How effective are the things people say to apologize? Effects of the realization of the apology speech act. *Journal of Psycholinguistic Research*, 26: 127–40.
- Schlenker, B.R. (1980). *Impression management: The self-concept, social identity, and interpersonal relations*. Monterey, CA: Brooks/Cole Publishing Company.
- Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk Analysis*, 13:675-682
- Spence, M. A. (1973). Job market signaling. *Quarterly Journal of Economics* 87( 3):355–374.
- Spence, M. A. (1974). *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Cambridge (MA): Harvard University Press.
- Stouten, J., & De Cremer, D. (2010). “Seeing is believing”: The effects of facial expressions of emotion and verbal communication in social dilemmas. *Journal of Behavioral Decision Making* 23:271-287
- Tavuchis, N. (1991). *Mea culpa: A sociology of apology and reconciliation*. Stanford, CA: Stanford University Press.
- Tomlinson, E. C., Dineen, B. R., & Lewicki, R. J. (2004). The road to reconciliation: Antecedents of victim willingness to reconcile following a broken promise. *Journal of Management*, 30: 165–187.
- Van Kleef, G.A., De Dreu, C.K.W., & Manstead, A.S.R. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology*, 86, 57–76.
- Van Kleef, G.A., De Dreu, C.K.W., & Manstead, A.S.R. (2006). Supplication and appeasement in conflict and negotiation: The interpersonal effects of disappointment, worry, guilt and regret. *Journal of Personality and Social Psychology*, 91, 124–142.
- Witvliet, C.V.O., Ludwig, T.E., & Bauer, D.J. (2002). Please forgive me: Transgressors’ emotions and physiology during imagery of seeking forgiveness and victim responses. *Journal of Psychology and Christianity*, 21, 219–233.
- Wubben, M.J.J., De Cremer, D., & van Dijk, E. (2009). When and how communicated guilt affects contributions in public good dilemmas. *Journal of Experimental Social Psychology* 45:15-23.
- Zahavi, A. (1975). Mate selection: a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214.

- Zahavi, A. (1977). The cost of honesty (further remarks on the handicap principle). *Journal of Theoretical Biology*, 67:603–605.
- Zahavi, A. (1993). The fallacy of conventional signalling. *Philosophical Transactions of the Royal Society London, Series B*, 340:227–230.
- Zahavi, A. & Zahavi, A. (1997). *The Handicap Principle: a Missing Piece of Darwin's Puzzle*. Oxford University Press: New York.
- Zheng, J., Veinott, E., Bos, N., Olson, J.S., & Olson, G.M. (2002). Trust without touch: jumpstarting long-distance trust with initial social activities. In *Proceedings of CHI*. ACM Press, New York, NY, pp. 141–146.

**Table 1: Game 1 Determinants of Investment and Return**

Dependent variable	<i>Invest1</i>	<i>Invest1</i>	<i>Return1</i>	<i>Return1</i>
Specification	(1)	(2)	(3)	(4)
	Probit	Probit	OLS	OLS
<i>Promise1</i>	0.05	0.98***	-0.02	3.04***
[promise in Game 1]	(0.04)	(0.18)	(0.11)	(0.46)
<i>Promise1sqr</i>		-0.04***		-0.15***
[promise squared]		(0.01)		(0.02)
<i>Constant</i>	0.50	-4.14***	8.35***	-6.88***
	(0.34)	(0.92)	(1.05)	(2.41)
Observations	229	229	191	191

\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%.

**Table 2: Game 2 Determinants of Investment and Return**

Dependent variable	<i>Invest2</i>	<i>Invest2</i>	<i>Return2</i>	<i>Return2</i>
Specification	(1)	(2)	(3)	(4)
	Probit	Probit	OLS	OLS
<i>Promise1</i>	0.01	-0.15	0.20	0.46
[promise in Game 1]	(0.07)	(0.25)	(0.13)	(0.51)
<i>Promise1sqr</i>		0.00		-0.01
[promise squared]		(0.01)		(0.02)
<i>Return1</i>	0.06	0.10	0.19	0.15
[return in Game 1]	(0.10)	(0.11)	(0.22)	(0.24)
<i>Broken1</i>	-0.49	-0.46	-0.17	-0.14
[broken promise in Game 1]	(0.43)	(0.45)	(1.09)	(1.11)
<i>Broken1</i> ×( <i>Promise1</i> - <i>Return1</i> )	-0.08	-0.03	-0.45*	-0.50*
[extent of broken promise]	(0.10)	(0.10)	(0.27)	(0.29)
<i>Distrusted1</i>	0.19	0.77	0.09	-0.16
[not trusted in Game 1]	(0.92)	(0.98)	(2.14)	(2.19)
<i>Promise2</i>	0.10**	0.56***	0.32***	-0.03
[promise in Game 2]	(0.04)	(0.15)	(0.12)	(0.44)
<i>Promise2sqr</i>		-0.02***		0.01
[promise squared]		(0.01)		(0.02)
<i>Constant</i>	-0.07	-1.89	1.98	2.80
	(0.75)	(1.26)	(1.93)	(3.04)
Observations	229	229	200	200

\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%.

**Table 3: Game 2 Determinants of Investment and Return with Game 1 Promise-Keepers**

Dependent variable	<i>Invest2</i>	<i>Invest2</i>	<i>Return2</i>	<i>Return2</i>
Specification	(1)	(2)	(3)	(4)
	Probit	Probit	OLS	OLS
<i>Return1</i>	0.30***	0.14	-0.13	-0.09
[return in Game 1]	(0.12)	(0.16)	(0.19)	(0.21)
<i>Promise2</i>	0.00	1.25***	0.97***	0.65
[promise in Game 2]	(0.07)	(0.36)	(0.16)	(0.71)
<i>Promise2sqr</i>		-0.05***		0.01
[promise squared]		(0.01)		(0.03)
<i>Message</i>	0.73**	0.62*	-0.31	-0.28
[message with content]	(0.35)	(0.38)	(0.63)	(0.63)
<i>Constant</i>	-1.69*	-7.16**	0.83	2.26
	(1.02)	(1.87)	(1.95)	(3.66)
Observations	155	155	143	143

\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%.

**Table 4: Game 2 Determinants of Investment and Return with Game 1 Promise-Breakers**

Dependent variable	<i>Invest2</i>	<i>Invest2</i>	<i>Return2</i>	<i>Return2</i>
Specification	(1)	(2)	(3)	(4)
	Probit	Probit	OLS	OLS
<i>Promise2-Promise1</i>	0.26**		-0.59	
[intended atonement]	(0.12)		(0.50)	
<i>Promise1-Return1</i>		-0.15**		-0.46*
[amount of broken promise]		(0.07)		(0.25)
<i>Promise2</i>	0.18	0.44	-5.03	-6.30
[promise in Game 2]	(0.54)	(0.53)	(4.01)	(3.71)
<i>Promise2sqr</i>	-0.01	-0.01	0.21	0.25*
[promise squared]	(0.02)	(0.02)	(0.15)	(0.14)
<i>Message</i>	2.09**	2.01**	16.49**	16.61**
[message with content]	(0.99)	(0.90)	(7.85)	(7.36)
<i>Apology</i>	0.18	0.50	-1.04	-0.91
[message with apology]	(0.69)	(0.73)	(2.18)	(2.08)
<i>Constant</i>	-1.91	-3.70	20.56	32.11
	(3.21)	(3.14)	(22.30)	(20.74)
Observations	36	36	25	25

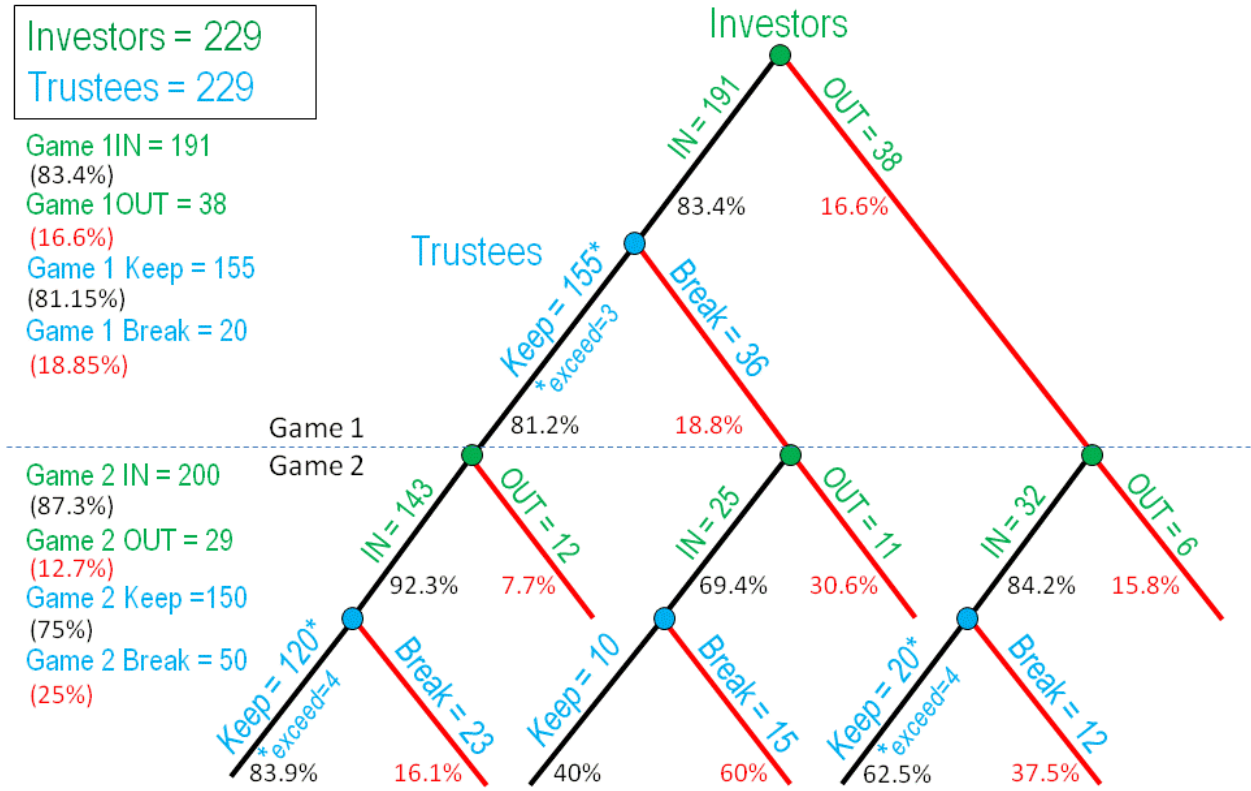
\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%.

**Table 5: Game 2 Determinants of Investment and Return with Game 1 Distrusted Trustees**

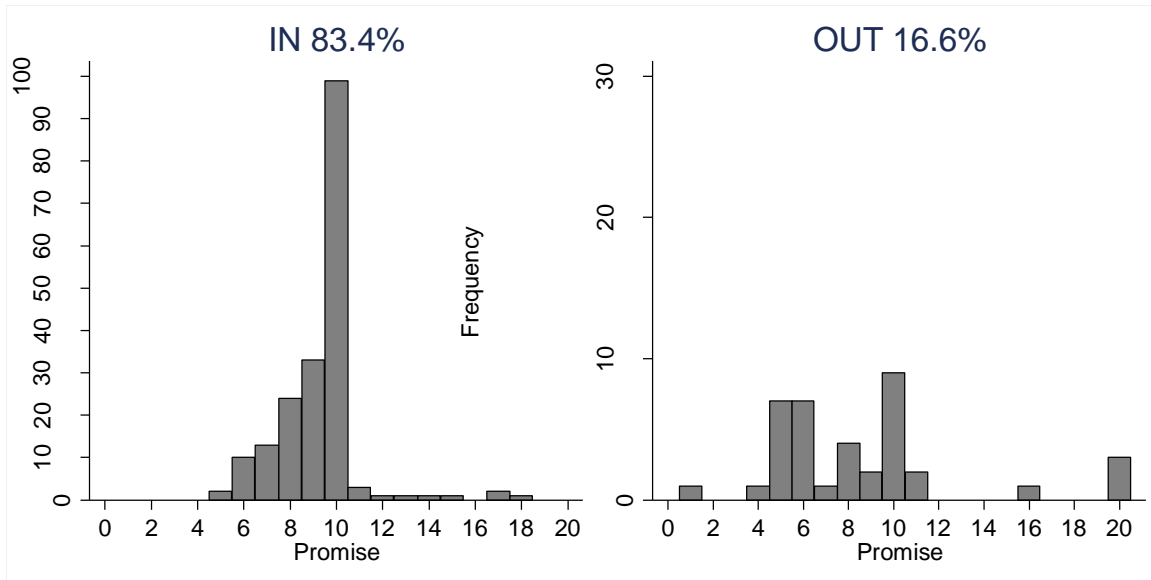
Dependent variable	<i>Invest2</i>	<i>Return2</i>
Specification	(1)	(3)
	Probit	OLS
<i>Promise2-Promise1</i>	0.03	-0.14
[intended atonement]	(0.09)	(0.21)
<i>Promise2</i>	-0.50	0.14
[promise in Game 2]	(0.43)	(0.73)
<i>Promise2sqr</i>	0.06*	0.01
[promise squared]	(0.04)	(0.03)
<i>Message</i>		1.89
[message with content]		(3.81)
<i>Wordcount</i>	0.03*	0.01
[number of words]	(0.02)	(0.04)
<i>Constant</i>	0.14	2.70
	(1.24)	(5.59)
Observations	36	32

\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%.

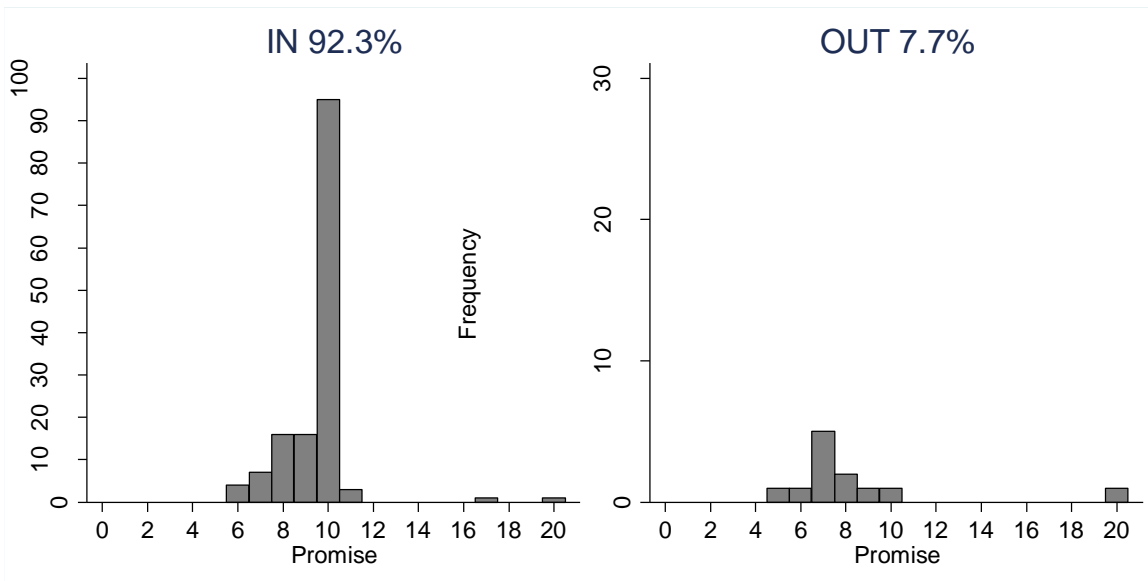
**Figure 1: Aggregate Distribution of Decisions in Games 1 and 2**



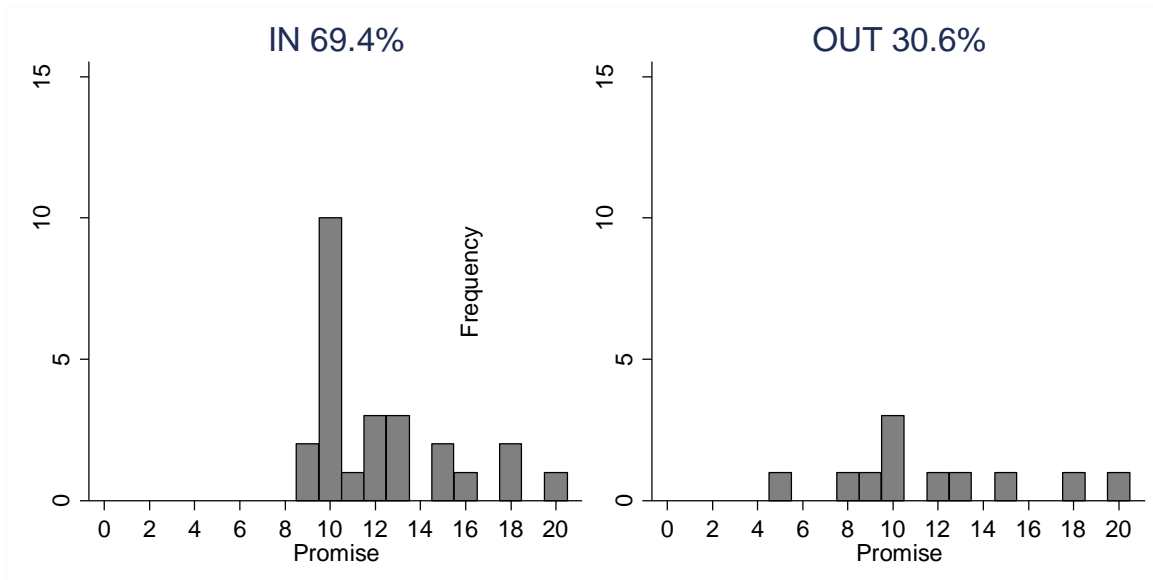
**Figure 2: Distribution of Promises in Game 1 (Resulting in IN or OUT)**



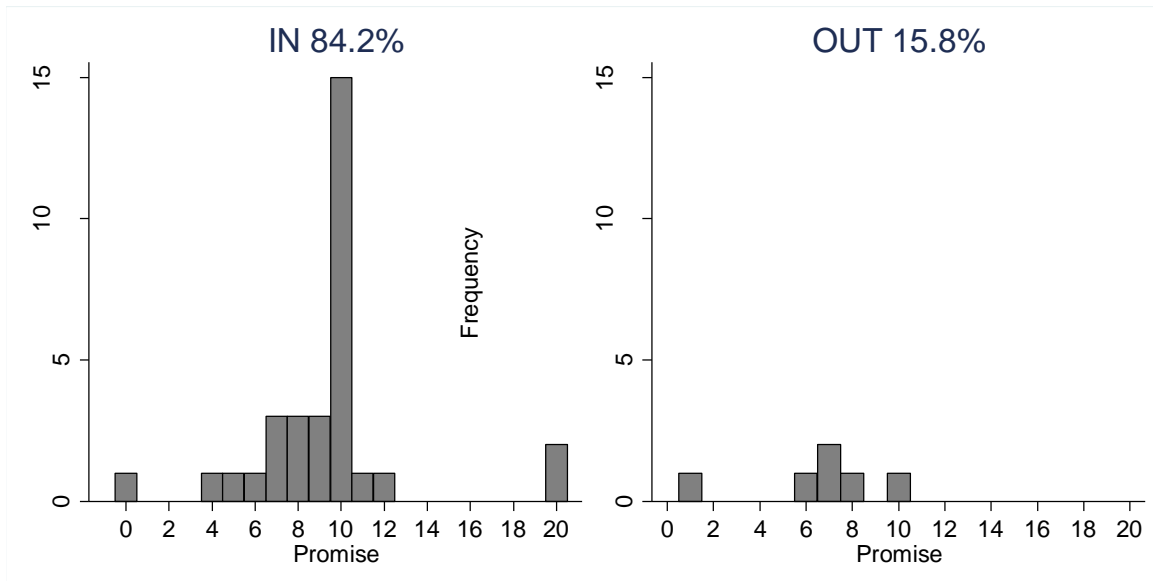
**Figure 3: Distribution of Promises in Game 2 (Resulting in IN or OUT) by Promise-Keepers**



**Figure 4: Distribution of Promises in the Second Game (Resulting in IN or OUT) by Promise-Breakers**



**Figure 5: Distribution of Promises in the Second Game (Resulting in IN or OUT) by Distrusted Trustees**





# Appendix A: Instructions

## INSTRUCTIONS

Thank you for participating in this experiment. The purpose of this experiment is to study how people make decisions in a particular situation. Feel free to ask us questions as they arise, by raising your hand. Please do not speak to other participants during the experiment. You will receive \$7 for participating in this session. You may also receive additional money, depending on the decisions made (as described below). Upon completion of the session, this additional amount will be paid to you individually and privately.

During the session, you will be paired with another person. However, no participant will ever know the identity of the person with whom he or she is paired.

## DECISION TASKS

In each pair, one person will have the role of A, and the other will have the role of B. The amount of money you earn depends on the decisions made in your pair.

First, by choosing a dollar amount from \$0 to \$20, B indicates the proportion of a possible \$20 income that he or she promises to transfer back to A, should A choose IN. Specifically, B will complete the following statement: "I (Participant B) promise to transfer back \_\_\_ of my income to you (Participant A) if you choose IN". The computer will convey B's statement to A, and then A and B will proceed as described below. B may still choose an amount to transfer back to A that is different than the amount promised.

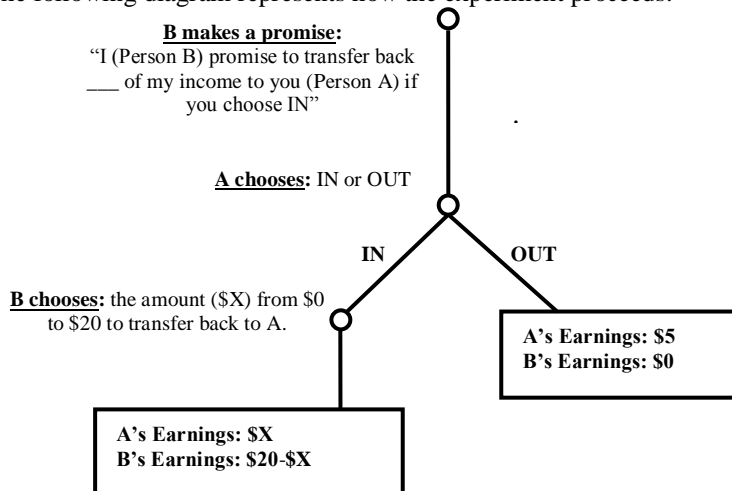
Having received a statement from B, A indicates whether he or she chooses IN or OUT. If A chooses OUT, A receives \$5 and B receives \$0. If A chooses IN, then B receives \$20 income. In such a case, after receiving \$20 income, B must choose a dollar amount from \$0 to \$20 to transfer back to A.

## SURVEY

After having completed the decision tasks described above you will be asked to fill out a short 20 item survey.

## DIAGRAM

The following diagram represents how the experiment proceeds:



*(This part of the instructions was handed out after the first part of the experiment was conducted.)*

#### REPETITION OF THE EXPERIMENT

The same decision tasks that were just completed will be repeated again, with everyone remaining in the same A or B roles and paired with the same participants as in the previous tasks.

#### MESSAGE

Prior to repetition of the previous decision tasks, B has an option to send a message to A. B may use a text box to type a message, if desired. We will allow time as needed to construct and type messages. When B's message has been completed (by typing in the text box and clicking on the send button) it will be conveyed by the computer to the appropriate Participant A, and then A and B will proceed with decision tasks. In these messages, no one is allowed to identify him or herself by name, number, gender, or appearance. Other than these restrictions, B may say anything in the message. If you wish not to send a message, simply click on the send button without having typed anything in the message box.

#### DECISION TASKS AND SURVEY (REPEATED AS BEFORE)

This second set of decision tasks and the accompanying 20 item survey is the final part of the experiment. There will be no further tasks.

## Appendix B: Messages

Table B1: Promise-Breakers Messages

Promised Game 1	Returned Game 1	Message	Word-count	Broad Apology	Narrow Apology	Promised Game 2	Trusted? Game 2	Returned Game 2
15	0	Let's split even. \$10 and \$10.		NO	NO	10	YES	8
10	1	If I knew there were 2 rounds I would have split it up even the first round. This round I'll make it up to you by giving you 15 if you're IN, this way we both end up with more money. Sorry again.		YES	YES	15	YES	3
10	0			NO	NO	20	YES	0
10	0	Hey im sorry about that I didn't realize there was going to be another round.! Let me make things right.		YES	YES	15	YES	5
10	9	i'll do the same deal as last time, sound fair?		NO	NO	15	YES	9
10	7	to even out i will give you 13 and i will take 7		YES	NO	13	YES	10
10	0	dooooooooood we all here to make muneey baby so why dont we just split this huney down da middle, a lil lovin for da both of us? ill forrealze give you like 10 bucks and ill keep 10 you dig? stay fresh :)		NO	NO	10	YES	0
10	7	How much would you want this time seeing how you didn't have a choice last time? I'm willing to make it even between the two of us.		YES	NO	13	YES	0
10	5	Hello A! I'm sorry I fell back on my promise, haha. To be honest, I'm dead broke and I haven't eaten all day and I'm literally about to run out of gas in my car, and those extra five dollars are going to help me out with that! ¶ If you choose out the most you are going to get is 5 more dollars, I can promise you that I'll agree to give you \$10 if you choose in. Hopefully this works out! Either way, have a good one!		YES	YES	12	YES	12
8	5	I only sent less than promised because I wanted to see what would happen		YES	NO	10	YES	10
10	2	Hi, I was a little confused as to the experiment before. But I will stay true to this promise		YES	NO	9	YES	9
13	9	lets split the money 10 and 10		NO	NO	10	YES	10
10	8	I apologize for cheating you out of your \$2 - normally I'm not the kind of person to do that sort of thing. When two people aren't face-to-face they usually have more confidence to do things they wouldn't normally do. This time I promise I'll play fair.		YES	YES	9	YES	9
10	7	I didn't know we were repeating this. This time I really will split 50/50 :)		YES	NO	10	YES	10
11	2	I will transfer back 18 to you this time to make it fair ¶ so we will have the same amount. I promise this time.		YES	NO	18	YES	10
14	11	Hi. I apologize for short changing you. I should have been honest and gone off the first example. I went off the third example w/my self interest in mind. I'll keep my word this time.		YES	YES	18	YES	18
11	10	10/10?		NO	NO	12	YES	7
8	0	sorry about last time i feel bad.....50/50 this time?		YES	YES	10	YES	0
8	7	Strategy :)		NO	NO	10	YES	10
10	0	I feel bad that you now only have the option of going home with \$5 so you should click in again and i will give you \$12 so that you go home with more than just the basic amount		YES	YES	12	YES	0

		possible						
8	6	My sincerest appologies on that last one...I do feel quite guilty ¶ and I assure you that this time I shall keep my promise with utmost integrity. You have my word as an honest gentleman.		YES	YES	16	YES	1
10	7	I feel bad for promising 10 and giving you 7. If you choose in I'll send you 13 so that we'll come out even.		YES	YES	13	YES	13
5	1	my bad...		YES	NO	10	YES	8
10	3	I'll transfer back more money this time ¶ actually \$10		YES	NO	11	YES	10
10	8	Hi ¶ I was testing if it really will let me decide how much I can get myself. This time I will give you the right amount I promised.		YES	NO	10	YES	10
10	0	This time I'll give you what I promise. Sorry!		YES	YES	10	NO	
10	0	ok for real this time haha. The first time was a joke lol		YES	NO	15	NO	
17	0	May God bless you		NO	NO	13	NO	
10	7	In the previous exercise I wanted to see if one really could promise one amount and then give another. After seeing that it is possible, I promise to give you the amount I state.		YES	NO	10	NO	
17	1	i know that there is no reason you'd trust me because i didn't follow through with my promise last time ¶ but if you choose in i will transfer all of the money that i say i will. ¶ for real this time.		YES	NO	18	NO	
10	5	Even though I was decietful ¶ you were no worse off then had you picked OUT. The other option would have still led you to \$5.		YES	NO	5	NO	
12	5			NO	NO	20	NO	
18	4			NO	NO	12	NO	
7	6			NO	NO	9	NO	
10	5			NO	NO	8	NO	
10	7			NO	NO	10	NO	

**Table B2: Distrusted Trustees Messages**

Promised Game 1	Message	Word-count	Promised Game 2	Trusted? Game 2	Returned Game 2
11	8 seems fair		8	YES	0
16	lets go 50/50. i give you \$10, i get \$10. ¶ its almost christmas....		10	YES	1
9	I want to split the money right down the middle. I will give you ten dollars and I will get ten dollars. If you choose out you will get less and both of us will come out empty handed. This is for the benefit of both parties and you will make more money in this way than you will by opting out.		10	YES	1
5	hey if i transfer 9 to you will you accept ?		9	YES	8
10	Trust me this time. Please?		10	YES	7
10	hi! i was actually going to give you the \$10 that time! You would have made more money! I promise to give what I promise to you this time as well!		8	YES	5
11	Hello, ¶ I think \$11.00 for you is a fair price for this survey and it is more than the \$5.00 you get for choosing Out. I will keep my offer the same if you chose In. ¶ Thank you		11	YES	12
10	Hey! Okay, listen, I was genuinely going to give you ten dollars. I think it makes sense for both of us to make as much money as possible. I'm not trying to trick you. I'm just poor and want a few extra dollars to buy Christmas presents. So could you please just be in" next time? That way we can both make more. I promise I am not lying to you. I know it's anonymous but please trust me. :("		9	YES	10
6	You click out, you earn 5. You accept my offer, you earn 6. It doesn't make any sense to click OUT. This is not a situation where my gain affects your profits in the future, this isn't one business earning a little bit and another earning a lot at its expense. You have to option of \$6 or \$5, without repercussions or any damage in the future. Me getting 0 does you no good, all it does is hurt you. If you want \$5, click OUT. But it obviously makes more sense to click IN.		6	YES	7
1	If I offer you at least 30% of my income we both make more than if you opt out.		7	YES	7
8	How about 10? We will both make the same amount evenly.		10	YES	10
10	I am a person of my word. I will transfer back \$10 so we both make the same amount of money and more money than if you pick OUT		10	YES	10
10	I will offer 10 dollars of my income to you. If you choose in, then you will receive 10 dollars and i will receive 10 dollars. If you choose out, you will only receive 5 dollars.		10	YES	10
20	I will split it with you so we both get ten dollars.		10	YES	10
6			9	YES	9
6	Please trust me when I say I will give you the amount I will promise you. This way, we will both earn more money instead of you just earning \$5 and me earning nothing. Let's take all of their money together!		7	YES	7
5	I will transfer 10 dollars.		10	YES	10
5	Hey ¶ to make this a win-win situation for both of us ¶ I'll transfer \$10 and that way both of us will earn the same amount. It's really a good gameplan. :)		0	YES	8
7	Ok ¶ so this time let's make it actually fair.... I should have made it even last time. So this time if I give you back \$8 ¶ you'll leave with \$20 and I'll leave with \$19. you still come out on top ¶ but I don't mind. And that's more than you'll make if you click OUT. I'm in the same boat as you....I too am poor as hell and would like to make some easy cash....		8	YES	8
20	I will transfer you back 75% back.		10	YES	0
5	We can figure out a way to divide the amount of the \$20 equally if the result from that will have us leave here with more than \$7		10	YES	9
10	I won't ask you to trust me. That's your choice ¶ what I will say though is offer you \$10 to each of us. We both walk away from this evenly and both better off than we came in.		10	YES	10
4	I promise to transfer back 20 of my income to you. I really need this extra money. I hope you understand		20	YES	0
9	Let's be fair and split the pool evenly. Trust that I will not go back on what I say.		10	YES	10
6	I will give you half of the amount of the income		12	YES	0
5	Hello A ¶ I'm stoked to be making money while my roommate snores away. Hahaha. ¶ Cha-ching ly ¶ B.¶		4	YES	4

20	if i say \$20 and you accpet ¶ I promise to give you \$20 back so we both leave with \$20 ¶ the max amount		20	YES	20
8	I promise to uphold any deals set before me		10	YES	10
6	I promise to give you \$7 for clicking "in." I guarantee it. As I see it ¶ this gives you \$2 more dollars than you would recieve by clicking "out." It's a win-win situation.		7	YES	7
8			10	YES	0
5	I think you should choose IN because it is simple game theory. If you choose IN and I choose to give you \$10 ¶ which I promise to give you ¶ then we both win. I know that you would automatically want to choose OUT so that you can get \$5 no matter what ¶ but I promise you that you will get \$10. We both want to get money ¶ and this is a good way to share our earnings. I hope you choose IN! :)		5	YES	0
10	If I offered you 10\$ why would you rather get 5?		10	YES	10
6	Ouch. ): I'm not gonna scam you, dude. When I make a promise, I make a promise. We both make more money this way; it's good all around!		6	NO	
5	hey Participant a make a deal dont do like this we should come here to earn money kul		1	NO	
10	dont be an asshole		7	NO	
10	Trust me.		10	NO	
6	You'll get more than \$5.		8	NO	
8	Hello there! So it's probably hard to trust me ¶ in that I will return your money? And I would quite frankly feel the same way. The thing is though that you don't know me but I know me and I know that when I make a promise I keep it. I hope you can trust in me. :)		7	NO	

**Table B3: Promise-Keepers Messages**

Promised	Returned	Message	Word-count	Promised Game 2	Trusted? Game 2	Returned Game 2
6	14	I paid out more than I promised to transfer back the first time as a reward for going IN		10	YES	0
6	6	merry christmas!		10	YES	10
10	10	Same deal as before sounds about right, in my opinion.		10	YES	10
8	8	i guess you need the money too so we should split it!		10	YES	10
10	10	Hey there. Want to do the same thing again, and both come out ahead?		10	YES	10
9	9	hey so 10 and 10 this time?		10	YES	10
10	10	I will split it equally		10	YES	10
10	10	Thanks for accepting my last offer. I promise to always uphold my side of the deal.		10	YES	10
9	9	hello A! :)		7	YES	7
7	7	I won't lie to you. I know we're all broke college students here who need to make money. ugh		8	YES	8
10	10	This is tres bizarre.		10	YES	10
5	5	i send you 10 and you hit in..that way we both get the same amount of money. =]		10	YES	10
10	10	Let's do the same thing, that way we both get the max amount of money		10	YES	0
10	10	we'll go 50/50 on everything. i promise.		10	YES	10
6	6	we're a good pair. i dont know what else to say haha.		6	YES	1
11	11	expecto patronum!		11	YES	2
10	10	Pleasure doing business with you :)		10	YES	10
9	9	:) I dont know what to say haha but ill split it 50 50 this time for you		10	YES	10
10	10	Let's make some MONEY :) click in on all of them and i'll try and make it as fair as possible.		10	YES	10
9	9	I hope you are satisfied with the amount of money I offered you. I will offer more this time.		10	YES	10
10	10	I don't really have anything to say...let's split the money 10-10 again		10	YES	10
10	10	\$10 is better than \$5. Trust me, I'm a doctor haha		10	YES	10
6	6			6	YES	6
7	7	I will do exactly the same thing as I did before.		7	YES	7
10	10	Lets split it 11/ 9 everytime, that way we both get more money IN than OUT? sound good? I don't think you can answer me. . .		9	YES	9
6	6	Again I will promise \$6. Please choose IN as it will maximize the profit that both of us can potentially made. I promise that I will send the full amount and if we can trust each other i will increase the amount I send in the following round. Thank you.		6	YES	6
10	10	Same as last time? It's only fair we earn the same amount.		10	YES	10
10	10	hi. i think it's best when we split it! makes it fair for everyone		10	YES	10
10	10	ill give u ten everytime if you choose IN then we both get ten dollars everytime we both go home with the same amount of money. again ten dollars a piece everytime go home with same amt. :)		10	YES	10
10	10			10	YES	10
6	6	want to choose in and then we take half? 10 each?		10	YES	10
8	8			8	YES	8
10	10	Let's keep going 50/50		10	YES	10
7	7	I promise to transfer you more money than last time.		9	YES	9
9	9	Hi, hope you're content with the \$9		10	YES	10
10	10	Let's split the 20 evenly, 10-10		10	YES	10
10	10	Want to just split it again?		10	YES	10
10	10	same thing as before, we both might as well walk out with enough for gas money!		10	YES	10
10	10	same thing?		10	YES	10
9	9			9	YES	9
10	10	I will keep it equal like last time.		10	YES	10
8	10			7	YES	9
9	9			9	YES	9
8	8	Same as before Ill send you 8. We both get more \$\$ that way!		8	YES	0
10	10	Same deal.		10	YES	10

10	10			10	YES	10
8	8			8	YES	5
10	10	i promise to do 50/50 again		10	YES	10
8	8			8	YES	8
8	8	hey, so i just want you to know that i'll probably sent you \$8 or \$9! nice working with you!		8	YES	8
10	10	I like the way we did it last time, it works out nicely for both of us and it's fair :) ¶ Thanks for being great!		10	YES	10
10	10	I will be fair.		10	YES	10
6	6	Hi A! :)		9	YES	6
8	8	I'm going to do the same thing.		8	YES	9
9	9	I hope you're having a great day!		8	YES	8
9	9	Teamwork!		10	YES	10
10	10	I promise not to screw you out of any money and to transfer back what I say I will. If you choose in¶ we'll both benefit more! =D¶		8	YES	8
10	10	Don't worry, we'll evenly split the money this time, too, just like last time. I won't try to scam you or anything, because that's below me. You'll get the 10 dollars that I promise you. :)		10	YES	10
10	10	I'm not a risk taker and I'm not a dick. I said I'd give back ten before, and I did. We both want money. You can make \$5 or \$10 because I will give you ten again. yayyy money=))		10	YES	10
10	10	If we do this again, i'm always going to keep it equal for both of us.		10	YES	10
10	10	Hi, so I know it's hard to trust someone who you don't even know but I'll be I'll do my best to make things work.		9	YES	9
10	10	I figure we are both equally desperate for cash.		10	YES	10
9	9	Hi Participant A ¶ I hope you trust me due to the previous round. I will take care of you and uphold to my promises, if you take care of me. Deal? Now lets do this and make some bank! ¶ Signed, ¶ Participant B		10	YES	10
8	8	I'm going to offer \$8 again. Hopefully you choose IN. That way we can both make a profit.		8	YES	8
10	10	Have you ever done this before?		10	YES	10
10	10			10	YES	10
10	10	Same thing? Seems fair? ...		10	YES	10
9	9	i promise i will give you what i say i will		10	YES	10
10	10	Thanks, glad we're both making a good amount of money! It's tough starting us off though! Wish you the best!		10	YES	5
10	10	Hello. Hope this doesn't sound creepy or anything. I think we should work together to get out of here with the same amount of money. I'm going to send over 10 again. :)		10	YES	10
10	10	I think each of us getting 10 dollars is fair. do you agree?		10	YES	10
8	8	same as last time :)		8	YES	8
9	9			9	YES	9
10	10	You can trust me :)		10	YES	10
10	10	Keep it even again		10	YES	10
10	10	i chose to give \$10 dollars and gave you \$10 in that last part. i hope we get paid		10	YES	10
10	10	I'm going to do the same thing as last time, 10 for you and 10 for me. We both would then walk away with 27 dollars :)		10	YES	10
8	8			8	YES	8
9	9			9	YES	9
10	10	Hope you like the wind...		10	YES	10
8	8	Were you happy with the outcome?		9	YES	9
10	10	Hey if you accept the \$10 then we both make that everytime and thats the most mutually beneficial.		10	YES	10
10	10	Same thing again. We both benefit.		10	YES	10
10	10	hi! let's split the money 50/50 and each get 10 every time		10	YES	10
7	7			7	YES	7
10	10	Thanks for choosing IN :) hopefully if we do the same thing again we'll both make \$20 each? thanks!		10	YES	10
10	10	Hello ¶ I wanted to make things 50/50. I don't really understand but that seemed fair to me at least		10	YES	10
7	7	I have no idea what to say here. This is a nice text box?		8	YES	1
9	9	I believe example 1 seemed the fairest for the position i was given. I did not want to be unfair however it seemed necessary to try and make a profit. I chose the smallest profit option which gave us both money in the		9	YES	9



		end.			
10	10	I'll give you \$10 just like before if you say "IN." ¶ It's a win-win (I get \$10 instead of \$0 and you get \$10 instead of \$5 if you were to say "OUT.")		10	YES 10
7	7	I need a nap...		11	YES 10
10	10	Let's just do that same transfer again		10	YES 10
8	8	=]		9	YES 9
10	10	Hi there ¶ just trying to keep things equal and honest ¶ now let's get some solid earnings again! :)		10	YES 10
8	8	This is a haiku. ¶ I am glad you trusted me¶This way we both win!		8	YES 10
7	7			8	YES 8
9	9	Please remember that if you say OUT ¶ you only get \$5. I PROMISE you that I will not give you under that if you say IN ¶ I promise.		7	YES 6
10	10	You're in good hands. Win/win.		10	YES 10
10	10	I want to keep this fair and even!		10	YES 10
10	10			10	YES 10
9	9	I'm glad you trusted me and went with IN ¶ I'm gonna do the same thing again so hopefully you go with IN again :)		9	YES 9
10	10	hi. i liked how we did it the first time. hopefully u did too		10	YES 0
10	10	I will send you 10 if you select IN ¶		10	YES 10
9	9	\$6 is the minimum offer to accept...anything higher your making more \$ just off generosity		10	YES 6
10	10			10	YES 10
9	9	Good deal! I'll up the transfer a to make it a litte more fair		10	YES 0
9	9	This time I'm going to promise 11 back to you ¶ and since you've seen I keep my promise ¶ when you click IN I will give you back 11 so we both walk out of here with 20. I don't break promises.		11	YES 11
9	9	50/50 ¶ sound good?		10	YES 10
10	10			10	YES 10
8	8	I'll give you exactly what I promise		10	YES 10
6	7			10	YES 8
10	10	just wanted to say hello :) have a nice day!		10	YES 10
10	10	Heyo- happy to work with you again ¶ and do the same thing.		10	YES 10
10	10	Hi! I'm going to split the money evenly. Have a nice day!		10	YES 10
9	9	Let's split it half and half ¶ ten dollars.		10	YES 10
10	10	Hello "A". Based on our last experiment we have established trust ¶ so thanks for making that happen! I will repeat the same steps as last time to ensure that we both get the same amount of money at our maximum level ¶ 10\$ each.it makes no sense to betray each other because we just come out of this thing with less money on both parts. Lets get rich!!		10	YES 1
10	10	im going to offer you 10 again ¶ take it and we can profit equally		10	YES 0
9	9	Hey just to let you know ¶ I try my best to never lie in life and I include this experiment part of my life standard so I won't lie.		6	YES 6
10	10	Hey. same amount ¶ same money ¶ we both leave with 27 buck in our pocket. =]		10	YES 10
10	10	Teamwork + Honesty = \$\$\$\$¶		8	YES 8
7	7	I Hate Mondays¶ -Garfield		17	YES 16
10	10			10	YES 10
8	8	Hey! So I want to make money ¶ just as much as you do ¶ so why dont we call it even and I promise \$10 ¶ you accept ¶ and we get out of here! =D Thanks		10	YES 10
10	10	Pay it forward. ¶ Have a great day.		10	YES 10
7	7			7	YES 7
10	10	Yay! great teamwork last time. I think we should do the same thing again this time. That way we both get the maximum amount of money. Hope that sounds good! :]		10	YES 10
10	10	hi hope your doing well. i plan on doing the same thing as before		10	YES 10
10	10	50-50 :]		10	YES 10
10	10	Same thing?		10	YES 0
10	10	I think we should do \$10 each again ¶ works out best for the both of us.		10	YES 10
9	9	Let's do the same...It worked and we both made some money!!!!		9	YES 9
10	10	Hey beautiful. I hopee your having a good day. Truthfully ¶ I'll get you more money if you say IN.		10	YES 8
9	9	well we worked together so far- want to do it again? at least we'll both make more than \$5		7	YES 7
10	10			10	YES 0

9	9	i will keep my promise!		9	YES	9
10	10	trust me		20	YES	20
7	7	I'm not quite sure what to say ¶ but hi!:) )		10	YES	0
9	9	lets do this!		10	YES	15
9	9			10	YES	10
9	9	same deal.		9	YES	9
10	10			10	YES	8
9	9			10	YES	10
9	9	I'm not entirely sure what I'm supposed to say ¶ BUT point is I promise I will not jip you out of money. What I promise is what you'll get and I hope you will not jip me out of any money either :)		9	YES	9
6	6			7	NO	
9	9			10	NO	
7	7	I'm planning on offering the same amount so we can potentially just do the same thing as before		7	NO	
8	8	choose IN ¶ i will transfer you the promised amount of \$		9	NO	
8	8			7	NO	
10	10	we need eachother to make money.		20	NO	
8	8	Hi		7	NO	
7	7			7	NO	
9	9			8	NO	
8	8	We the People of the United States of America, ¶ Inorder to form a more perfect Union, ¶ Do ordain and establish this constitution of the United States...		8	NO	
10	10			5	NO	
6	6	I'll promise to transfer whatever amount I say		6	NO	