

# MPRA

Munich Personal RePEc Archive

## **Building and Rebuilding Trust with Promises and Apologies**

Schniter, Eric and Sheremeta, Roman and Sznycer, Daniel

2013

Online at <https://mpra.ub.uni-muenchen.de/53596/>

MPRA Paper No. 53596, posted 10 Feb 2014 15:14 UTC

# Building and Rebuilding Trust with Promises and Apologies

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

September 9, 2012

## Abstract

Using trust games, we study how promises and messages are used to build new trust where it did not previously exist and to rebuild damaged trust. 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 design allowed us to observe the endogenous emergence and natural distribution 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 long messages and promises closer to equal splits to encourage trust in the second game. To restore damaged trust, promise-breakers used apologies and upgraded promises. On average, investments in each game paid off for investors and trustees, suggesting that effective use of cheap signals fosters profitable trust-based exchange in these economies.

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

---

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

\* The earlier version of this paper circulated under the title “Restoring Damaged Trust with Promises, Atonement and Apology.” For inspiration to pursue this study we thank John Dickhaut. We thank an advisory editor and an anonymous referee for their comments. Helpful comments were also received from Hilly Kaplan, Wojtek Przepiorka, and participants at the Workshop on Communication in Games (at the University of Zurich), the Human Behavior and Evolution Society annual meeting (in Montpellier, France), the Center for Evolutionary Psychology (at UC Santa Barbara), the John Dickhaut Memorial Conference (at Chapman University), and the Association for Religion, Economics and Culture annual meeting (at Chapman University). We would also like to thank the Economic Science Institute at Chapman University for funding this research.

## 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 encourage trust where it did not previously exist and restore trust when it is damaged 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 new trust can be encouraged where it did not previously exist and how damaged trust can be rebuilt (Dirks et al. 2009). Most of the existing research in this area (but see Fischbacher & Utikal 2010) is either exclusively theoretical (Lewicki & Bunker 1996; Mishra 1996; Lewicki & Wiethoff 2000; Ren & Gray 2009; Gillespie & Dietz 2009), based on anecdotal or archival evidence (Elsbach 1994; Knight & Pretty 1999), surveys (Slovic 1993), diary studies (Conway & Briner 2002) fictional vignettes (Tomlinson et al. 2004), videotaped dramatizations (Kim et al. 2004, 2006), or 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 study wherein financially motivated participants used endogenously created and naturally distributed promises and apologies. Our study 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

---

<sup>1</sup> This interpretation is based on the assumption that participants identify and act in accordance with unstated *if-then* propositions and expect others to do as well (Rousseau 1989), though there is no contract stating expected or contingent behavior in the classic “investment game” (see Berg et al. 1995). Because the assertion that the original game was universally understood to be 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.”

reciprocity in an investment setting (for a 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. 2006; 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.

In this paper we describe a study using trust games that examines how promises and messages are used to build new trust where it did not previously exist and to rebuild damaged trust. 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 design allowed us to observe the endogenous emergence and natural distribution 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 – compared to previously trusted promise-keepers – relatively longer messages to encourage new trust in the second game. Promise-breakers used relatively higher new promises (compared to all other trustees) and messages (usually with apology) to successfully restore damaged trust. On average, investments in each game paid off for investors and trustees, suggesting that the context-specific signaling described above, can foster profitable trust-based exchanges in these economies.

## **2. Background**

While mutually beneficial non-binding agreements help realize opportunities to gain from asynchronous trade, they are subject to exploitation by under-reciprocators or non-reciprocators. Our research focuses on trustees’ cue and signal effects on investor trust in asynchronous exchanges that provide opportunity for mutual advantage. In these exchanges, we define trust as

voluntarily ceding resources to another in the expectation that the other intends to reciprocate in accordance with signaled intentions. Trustworthiness is defined as reciprocation (of resources ceded by the investor) in accordance with signaled intentions.

To successfully navigate a trust-based cooperative interaction and avoid exploitation by cheaters, it is important for investors to obtain accurate information about the ability and willingness (propensity) of trustees to carry out their end of the cooperative deal. Trustworthy reputations that have been demonstrated by past actions serve as reliable *cues* upon which investors can make trust-based decisions. In the initial interactions with unknown partners, informative cues about an investor's willingness to trust or a trustee's trustworthiness are unavailable. In the absence of information about the interactants' past behavior, *signals*<sup>2</sup> are often sent to receivers with the intention to communicate information about the sender (e.g., see Farrel & Rabin 1996); for example, that the sender is trustworthy. Where cues have informed investors of untrustworthiness, signals may be sent with the intention of persuading those investors that the sender is more trustworthy than inferred from those cues alone.

Signals encouraging trust appear to be important tools for developing mutually beneficial relationships under conditions where trust has not yet been established and where trust has been damaged. Without the effective use of signals cooperative interactions may be foregone: potential investors may decide not to extend trust when they lack reputational information and when cues indicate a breach of trust. Further, when trust has been damaged, signals give investors access to relevant though invisible propensities of trustees, such as in the case of recalibrated upgrades in trustworthiness. This is true whether trust has been damaged intentionally or unintentionally (Axelrod & Dion 1988).

Although signals that accurately convey behavioral propensities are potentially useful to both senders and receivers, signalers may send "dishonest signals" to benefit at the expense of receivers. Critical receivers can incur lower costs than naïve receivers (Dawkins & Krebs 1978; Maynard Smith 1982), and so natural selection favors those receivers who can accurately assess

---

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

the cost-benefit tradeoffs associated with emitters' signals, and calibrate their trustfulness accordingly.

Zahavi (1975) addressed the question of "why are signals reliable?" suggesting that the high production cost of a signal guarantees its reliability, insofar as the production cost outweighs the benefits gained from using the signal deceptively, but not from using it honestly. The prototypical example is the massive and colorful peacock's tail, indexing the peacock's genetic quality for peahens' mate selection (Petrie et al. 1991). In this system, costly signals persuade the receivers while cheap signals fail to do so (Zahavi 1993; Grafen 1990).

Production costs are not the only warrantors of signal reliability, however. Human language, whether spoken or written, is an arbitrary communication system that often uses relatively cheap-to-produce signals to negotiate trust between individuals with conflicting interests (Lachmann et al. 2001). If these cheap signals are used by and relied upon by humans globally and on average, what explains the maintenance of their reliability?

The reliability of cheap signals is supported by the actuality or threat of social sanctions that can more than offset the short-term benefits of cheating and deception (Rohwer 1977; Kiyonari et al. 2000; Masclet et al. 2003). A selective regime characterized by repeated interactions among known others (Kelly 1995) has led to psychological mechanisms for social exchange that balance (i) the costs of mistaking a one-shot interaction for a repeated interaction with (ii) the far greater costs of mistaking a repeated interaction for a one-shot interaction (Delton et al. 2011). Hence, participants in explicitly one-shot anonymous experiments often behave as if they expect repeated interactions with trustworthy, intrinsically valuable partners (e.g., see Hoffman et al. 1996; Kiyonari et al. 2000).

While the sanctioning of false signals and our tendency to err to caution may reduce the frequency of false signals in a population, those individuals who expect to escape sanctions may be more motivated to use signals deceptively. In economies where opportunity costs of forgone trust-based exchange are larger, receivers tend to tolerate greater proportions of false signals to honest signals. 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 (i.e., when the consequent exchange

produces a loss). This economically justified tolerance of a rate of false signaling also predicts that individuals will exploit opportunities to use deception. This study explores the use of cheap signals (e.g., promises of reciprocation, personalized messages, and apologies) that do not directly affect payoffs of the game, or require monetary costs for production, yet are common features of trust-based interactions. Personalized communication may improve cooperation (Orbell et al. 1988; Bohnet & Frey 1999; Ridings et al. 2002; Zheng et al. 2002; Buchan et al. 2006) by facilitating coordination, decreasing social distance, raising solidarity, and providing the cues of familiarity that are normally associated with trustworthy relationships. Non-binding promises have also been shown to increase cooperation (Rubin & Brown 1975; Kerr & Kaufman-Gilliland 1994; Elingsen & Johannesson 2004; Charness & Dufwenberg 2006). In relationships where trust has been damaged, apologies and explanations have been shown to elicit forgiveness (Ohbuchi et al. 1989; Tavuchis 1991; Lewicki & Bunker 1996; Benoit & Drew 1997; Girard & Mullet 1997; McCullough et al. 1997, 1998; Girard et al. 2002; Witvliet et al. 2002) and promote future trust (De Cremer et al. 2010). These strategies are based on signals that are cheap to produce, raising the questions of how people use signals in these contexts; when the signals achieve their intended effects; and who benefits from their use.

In sum, while cheap signals are helpful for building new trust and rebuilding damaged trust to achieve efficient outcomes, they can be used deceptively and may be distrusted, making their reliability tenuous. Therefore, we expect that in our experiment trustees whose actions have already produced reliable cues establishing their trustworthy reputations (by keeping promises and not succumbing to more profitable opportunism) will be less incentivized (than previously distrusted trustees, or trustees whose reputations indicate untrustworthiness) to spend time and effort constructing messages to persuade investors to trust them, when those messages might be distrusted. Previously distrusted trustees who have not established trustworthiness and untrustworthy trustees (i.e., promise-breakers) are expected to make use of promises and messages in an attempt to affect investors' decisions to trust. We also 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.

### 3. 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 and did not involve deception, 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 (i) no indication of a subsequent game to follow and (ii) no promises that the experiment would end at conclusion of that game.

Participants were assigned to one of two roles: "Participant A" (investor), or "Participant B" (trustee). First, the trustee completed the following standardized statement (which we will refer to below as a *promise*) by selecting a natural number amount from 0 to 20: "I (Participant B) promise to transfer back \$\_\_\_ of my income to you (Participant A) if you choose IN". This statement was not binding, however. That is, the trustee was *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 OUT or IN. If the investor chose OUT, she received \$5 and the trustee \$0. If the investor chose IN, then the trustee received \$20 (the "income"), after which he selected a whole dollar amount from \$0 to \$20 to send back to the investor.

After the first trust game (Game 1) finished, participants were given instructions (see Appendix A) indicating that a second trust game (Game 2) would follow. In Game 2, participants were told they would remain in the same roles and interact with the same partner as in Game 1. However, prior to Game 2, the trustee was given an opportunity to use a text box to send a one-way message to the investor. Trustees were told 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 subsequently the standardized promise to the investor, and then Game 2 proceeded. We specified that Game 2,



which had the same rules as Game 1, was the last and final part of the experiment (i.e., there would be no subsequent games).<sup>3</sup>

There were 25 experimental sessions. Each session had between 10 and 24 participants. The average experimental earnings were \$18, ranging from a \$0 to \$40, plus \$7 for arriving to the experiment on time and participating. No participant participated more than once, and no participant had prior experience with a similar game environment.

## 4. Results

### 4.1. Game 1

We expect that trustees, aware of investor self-interest and motives for critical signal reception, would promise investors transfers of at least \$6 (minimally higher than the payoff to the investor if he chooses OUT) but less than \$20 (which would provide no benefit to the promise-maker). Two plausible focal points for promised return amounts are the midpoint of the \$6-\$19 range, \$12.5 (though only whole dollar amounts like \$12 or \$13 could be chosen), and the even-split of \$10. Wary that trustees' may have less incentive to honor promises closer to \$20 than to the even-split amount of \$10, we also expect that investors should be more suspicious of the veracity of higher promises and therefore less likely to invest in higher promises. If the mind errs to caution, as we have suggested, and interprets the one-shot game as potentially repeatable, trustees who have been trusted should reciprocate enough to, at minimum, provide investors profitable returns on their investments. These predictions stand in stark contrast to the rational (non-cooperative) 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.

---

<sup>3</sup> After each trust game participants 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.

First we evaluate the distributions of Game 1 promises associated with trusting and distrusting decisions, respectively, and how these investment decisions affected investor and trustee earnings. The distribution of promises in Figure 2 indicates that investors who chose IN received promises in the range of \$6-\$19 (99% of the time), with promises of \$12 or \$13 relatively uncommon (1% of the time), and the promise of \$10 most common (more than 50% of the time). Investors who chose OUT received lower promises on average (i.e.,  $M_1=\$8.61$  ( $SD_1=4.33$ ) versus  $M_2=\$9.31$  ( $SD_2=1.75$ ); Wilcoxon rank-sum test,  $p\text{-value}=0.01$ ,  $n_1=191$ ,  $n_2=38$ ), and, compared to trusted promises, received either relatively low or relatively high promises overall. To confirm this observation, we estimate probit models (see Table 1, specifications 1 and 2), where the dependent variable is the investment decision in Game 1 and the independent variables are dummy variables for promises less than \$9 and greater than \$11, as well as the amounts of these promises. In specification (1), the dummy variables are negative and significant, indicating that investors are less likely to invest when promises are either relatively low or relatively high. Moreover, specification (2) indicates that, among promises lower than \$9, there is a positive correlation between the amount promised and the probability of investment. On the other hand, among promises higher than \$11, there is a negative correlation between the amount promised and the probability of investment. In other words, promises closer to the even-split of \$10 elicit a higher rate of IN responses.

In Game 1, investment yielded greater payoffs than non-investment for both investors and trustees. Investors who chose IN received back \$8.19 on average, which is more than their original endowment of \$5 (Wilcoxon signed rank test,  $p\text{-value}<0.01$ ,  $n=191$ ). Trusted trustees earned an average of \$11.81, more than the \$0 of distrusted trustees (Wilcoxon rank-sum test,  $p\text{-value}<0.01$ ,  $n_1=38$ ,  $n_2=191$ ). The OLS estimation of specifications (3) and (4) in Table 1 indicates that the amount returned by trustee is correlated with the amount promised. Specifically, specification (3) indicates that both low and high promises are followed by lower returned amounts. Moreover, specification (4) indicates that amount returned increases as it gets closer to the even-split promise of \$10. These results support the rationale for why investors receiving especially high or low promises should tend to choose OUT.

For the investors who chose IN, the mean amount returned of \$8.19 was significantly lower than the mean trusted promise of \$9.31 (Wilcoxon signed rank test,  $p\text{-value}<0.01$ ,  $n_1=n_2=191$ ). Despite mean returns being lower than promised, we find that promises 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 use the terms “promise-keepers” and “promise-breakers” to refer to trusted trustees who exactly met *or* exceeded their promised amounts, and who returned less than their promised amounts (whether the returns were monetarily profitable to the investors or not), respectively.

## 4.2. 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 many investors and trustees – raising the question of whether the use of cheap signals will still matter where cues of trusting and trustworthy behavior (or its absence) are available.

In Game 2, trustees promised to return \$9.79 on average, a higher amount than the mean of \$9.20 promised in Game 1 (Wilcoxon signed rank test,  $p$ -value $<0.01$ ,  $n=229$ ). Game 2 promises 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,  $n=229$ ). Trustee reputation as established in Game 1 and the new promises issued in Game 2 affected investment decisions in Game 2. The estimation of probit models (specifications 1 and 2 in Table 2)<sup>4</sup> indicates that both promise-breaking in Game 1 and promises lower than the even-split in Game 2 elicited less investment decisions in Game 2. Overall, the investments made in Game 2 paid off. Investors who chose IN received back \$8.73 on average, which is more than the OUT payoff of \$5 (Wilcoxon signed rank test,  $p$ -value $<0.01$ ,  $n=200$ ). Trusted trustees earned an average of \$11.27, more than the \$0 of distrusted trustees (Wilcoxon rank-sum test,  $p$ -value $<0.01$ ,  $n_1=29$ ,  $n_2=200$ ). The estimation of specifications (3) and (4) in Table 2 indicates that higher promises in Game 1, a greater extent of under-return relative to promise in Game 1, and uneven split promises in Game 2 all predict lower amounts returned in Game 2. Overall, and similar to Game 1, promises in Game 2 tended to be veridical; 75% of promises (150/200) were

---

<sup>4</sup> In estimation of Table 2, we have excluded variables corresponding to the amount promised in Game 1 (i.e.,  $Promise1 < 9 \times Promise1$  and  $Promise1 > 11 \times Promise1$ ), since these variables are highly correlated with returns. Nevertheless, even when these variables are included the estimates in Table 2 are very similar.

kept or exceeded, and 25.0% (50/200) were broken. In the sections below we further explore the effects of promises and messages on Game 2 investments and earnings across different “types” of dyads aggregated by Game 1 decisions.

#### 4.2.1. Promise-Keepers

For the subset of 155 Game 1 promise-keeping trustees, we observe higher average promises in Game 2. Figure 3 displays the distribution of Game 2 promises resulting in IN and OUT made by these promise-keepers. Overall, this set of 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 promise-keepers’ demonstrated trustworthiness in Game 1, the Game 2 investment rate of 92.3% (143/155) in Game 1 promise-keepers was higher than the overall investment rate of 83.4% (191/229) in Game 1 (Fisher’s exact test,  $p\text{-value} < 0.01$ ,  $n_1 = 155$ ,  $n_2 = 229$ ). Specifications (1) and (2) in Table 3 indicate that messages with content and promises closer to the even split of \$10 positively affected investment in promise-keepers.<sup>5</sup>

We expected that, due to the greater reliability of available behavioral cues which demonstrated their trustworthiness (relative to the reliability of a cheap signal), Game 1 promise-keepers would be less inclined than promise-breakers and distrusted trustees to construct messages for the purpose of persuading investors to choose IN in Game 2, and so would send both shorter messages and a higher proportion of empty messages in Game 2. Supporting that expectation, Game 1 promise-keepers’ messages contained fewer words than those from the set including both Game 1 distrusted trustees and Game 1 promise-breakers ( $M_1 = 11.41$  ( $SD_1 = 11.94$ ) versus  $M_2 = 22.9$  ( $SD_2 = 22.37$ ); Wilcoxon rank-sum test,  $p < 0.01$ ,  $n_1 = 155$ ,  $n_2 = 74$ ). Game 1 promise-keepers’ messages were also more frequently empty than those from the set of both Game 1 distrusted trustees and Game 1 promise-breakers (20.6% versus 10.8% of the time; Fisher’s exact test,  $p\text{-value} = 0.05$ ,  $n = 229$ ).

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

---

<sup>5</sup> Note that in estimating these regressions we cannot include both returns in Game 1 and promises in Game 1 since for promise-keepers they are perfectly correlated. Moreover, we had to omit variable  $Promise2 > 11 \times Promise2$  since there are only three observations greater than \$11, which makes  $Promise2 > 11$  and  $Promise2 > 11 \times Promise2$  almost perfectly correlated.

OUT (Wilcoxon signed rank test,  $p$ -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 those Game 1 promise-keepers that were not trusted in Game 2, promise-keepers also profited from trusted promises in Game 2, earning \$11.38 on average (Wilcoxon rank-sum test,  $p$ -value $<0.01$ ,  $n_1=12$ ,  $n_2=143$ ). Specifications (3) and (4) in Table 3 indicate that Game 2 promises lower and higher than the even-split are associated with lower and higher amounts returned by Game 1 promise-keepers, respectively. Whether the messages have content or not, on the other hand, has no effect on returns.

#### 4.2.2. Promise-Breakers

A major question our data address concerns what happens after a breach of trust when a fresh opportunity for cooperation arises: How trustees behave, how investors respond, and what outcomes are achieved. Here we focus on the 18.8% (36/191) of pairs with broken promises in Game 1 (i.e., where the amount returned was lower than the amount promised). These broken promises represent breaches of trust and the relationships that immediately follow are considered to have *damaged trust* (because trust-based expectations were not met). A central question motivating this study is whether signals such as new promises and apologies can (i) restore investors' willingness to trust and (ii) facilitate the achievement of higher joint payoffs.

Figure 4 displays the distribution of promises made by 36 Game 1 promise-breakers in Game 2 resulting in IN and OUT decisions. Promise-breakers promised \$12.11 in Game 2, which is significantly higher than their average promise of \$10.58 in Game 1 (Wilcoxon signed rank test,  $p$ -value=0.01,  $n_1=n_2=36$ ). The extent of upgraded promises (*Promise2-Promise1*) by promise-breakers is also significantly higher than the extent of upgraded promises by promise-keepers ( $M_1=1.53$  ( $SD_1=3.70$ ) versus  $M_2=0.44$  ( $SD_2=1.79$ ); Wilcoxon rank-sum test,  $p$ -value $<0.01$ ,  $n_1=36$ ,  $n_2=155$ ). Assuming that many of the investors whose trust had previously been damaged would be inclined to choose OUT, it appears that promise upgrades partially restore trust, since 69.4% (25/36) of investors whose trust was damaged in Game 1 chose IN again.

In addition to promise upgrades, we also find that Game 1 promise-breakers frequently used messages. Table C1 in Appendix C 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.<sup>6</sup> Game 1 promise-breakers' messages contain more words than messages from Game 1 promise-keepers ( $M_1=19.06$  ( $SD_1=19.03$ ) versus  $M_2=11.41$  ( $SD_2=11.94$ ); Wilcoxon rank-sum test,  $p\text{-value}=0.03$ ,  $n_1=36$ ,  $n_2=155$ ), suggesting that the behavioral cue of trustworthiness made verbal persuasion for continued investment less determinant of re-investment and less necessary for trustees.

To further classify all 36 messages, we used an incentivized laboratory coordination game (Houser & Xiao 2011). Three coders recruited from the subject pool and blind to the hypotheses<sup>7</sup> were asked to code each message twice: first based on whether or not it conformed to a "broad" definition of apology (an explicit or implicit acknowledgment of offense), and second based on whether or not it conformed to a "narrow" definition of apology (an explicit or implicit acknowledgment of offense, along with remorse, regret, or sorrow stemming from acknowledgment of the offense). All 6 messages without content were coded by all coders as not conforming to the broad definition and not conforming to the narrow definition of apology. Of the 30 messages with content, 28 were coded by the majority of coders as conforming to the broad definition of apology and 13 were coded by the majority of coders as conforming to the narrow definition of apology.<sup>8</sup>

When using a broad definition of apology, which was coded with "substantial" agreement (Kappa of 0.70), we find that 82.1% (23/28) of apologizers were retrusted in Game 2 in comparison to only 25.0% (2/8) of non-apologizers (Fisher's exact test,  $p\text{-value}<0.01$ ,  $n=36$ ), suggesting that messages with apology are more likely to restore trust after broken promises than empty messages or messages without apology.<sup>9</sup> Due to the lower interrater reliability for messages coded according to the narrow definition, we will consider only the broad definition of apology in the subsequent analyses.

---

<sup>6</sup> We find that 80% (24/30) of messages with content restored trust (i.e., where investors chose IN in Game 2 after having suffered 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,  $p\text{-value}<0.01$ ,  $n=36$ ).

<sup>7</sup> The instructions for coders and details about how they were paid are attached in Appendix B. Coders each earned an average of \$28.33 for matched codings, plus \$7 for arriving on time and participating.

<sup>8</sup> We use a standard approach from content analysis methodology to calculate Cohen's Kappa interrater agreement coefficient (Cohen 1960; Krippendorff 2004). Kappa values between 0.41 and 0.60 are considered "Moderate" agreement, and those above 0.60 indicate "Substantial" agreement (Landis & Koch 1977). We find Kappa values of 0.70 and 0.53 for the broad and narrow definitions of apology, respectively.

<sup>9</sup> When using a narrow definition of apology, which was coded with moderate agreement (Kappa of 0.53), we find that 84.6% (11/13) of apologizers were retrusted in comparison to only 60.9% (14/23) of non-apologizers (Fisher's exact test,  $p\text{-value}=0.13$ ,  $n=36$ ).

Thus far, we have only considered the independent effects of new promises and apologies in restoring damaged trust, but recognize that these remedial strategies are often used jointly. Among Game 1 promise-breakers, the size of the upgrade in amount promised is significantly larger for participants issuing apologies than for those who did not ( $M_1=\$1.68$  ( $SD_1=3.10$ ) versus  $M_2=\$1$  ( $SD_2=5.55$ ); Wilcoxon rank-sum test,  $p\text{-value}=0.05$ ,  $n_1=28$ ,  $n_2=10$ ). When the apologetic promise-breakers are compared to all other trustees the difference is even larger. The upgrade in amount promised for apologetic trustees is almost four times greater than among all other trustees ( $M_1=\$1.68$  ( $SD_1=3.10$ ) versus  $M_2=\$0.44$  ( $SD_2=2.68$ ); Wilcoxon rank-sum test,  $p\text{-value}=0.05$ ,  $n_1=28$ ,  $n_2=201$ ), indicating that apologetic trustees upgraded their promises most.

Next, we estimate probit regressions (see 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 are promise adjustments (specifically promise upgrades) and apologies. Specification (2) shows that in addition trust is negatively affected by the extent of under-return relative to Game 1 promise. Moreover, based on the likelihood-ratio test, we find that the promise adjustment, upgraded relative to the extent of under-return on previously broken promise (i.e.,  $Promise2-(Promise1-Return1)$ ), positively and significantly influences trust in Game 2 (likelihood-ratio test,  $p\text{-value}=0.05$ ). These results indicate that investors' decisions to re-invest are sensitive not just to the existence of broken promises (specifications 1 and 2 in Table 2), but also to the extent of under-return relative to Game 1 promise, apologies, and upgraded promises (specifications 1 and 2 in Table 4).

We have argued that signals such as apologies and promises should have evolved only if they provided net benefits to both the senders and receivers of those signals on average and over the evolution of the communication system. 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 \$7.28 on average, which is significantly higher than the OUT payoff of \$5 (Wilcoxon signed rank test,  $p\text{-value}=0.05$ ,  $n=25$ ). Moreover, Game 1 promise-breakers returned significantly more in Game 2 than in Game 1 ( $M_1=\$7.28$  ( $SD_1=4.86$ ) versus  $M_2=\$4.60$  ( $SD_2=3.72$ ); Wilcoxon signed rank test,  $p\text{-value}<0.01$ ,  $n_1=n_2=25$ ). This is also true when we look at investments in the subset of 23 out of 28 trustees who issued apologies and where re-trusted: they returned significantly more in the second game ( $M_1=\$7.52$  ( $SD_1=4.81$ ) versus  $M_2=\$4.61$

( $SD_2=3.64$ ); Wilcoxon signed rank test,  $p\text{-value}<0.01$ ,  $n=23$ ), which is also significantly higher than the OUT payoff ( $M_1=\$7.52$  ( $SD_1=4.81$ ) versus  $M_2=\$5$ ; Wilcoxon signed rank test,  $p\text{-value}=0.03$ ,  $n=23$ ).

Although investments in Game 2 paid off, we still find that 60.0% (15/25) of re-trusted promise-breakers subsequently broke their promises again in Game 2 – almost irrespective of the apologies and new adjusted promises. From specifications (3) and (4) in Table 4, it appears that neither promises adjustments, new amounts promised, nor apologies are predictive of return in Game 2. The only variable that predicts return in Game 2 is return in Game 1.

### 4.2.3. Distrusted

As mentioned above, 16.6% of trustees (38 out of 229) were not trusted in Game 1 (see Figure 1). The source of this distrust appears to be: (i) the higher variance around the even-split point of the distribution of distrusted promises (relative to that of the distribution of trusted promises; see right panel of Figure 2), and (ii) a lower degree of default trustfulness among Game 1 distrustful investor (accounting for the fact that a sizeable number of even-split promises were rejected). 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 previously distrusted trustees would adjust their Game 2 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 trustee.

First we evaluate whether Game 1 distrusted trustees adjust 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,  $p\text{-value}=0.45$ ,  $n_1=n_2=38$ ), 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 adjusted promises. We expect that distrusted trustees would construct longer messages (to persuade investors to choose IN in Game 2) than trustees who had already established reputations of trustworthiness. Table C2 in Appendix C 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. Messages from Game 1 distrusted trustees contain more words than messages from Game 1 promise-keepers ( $M_1=26.58$  ( $SD_1=24.83$ ) versus  $M_2=11.41$  ( $SD_2=11.94$ ); Wilcoxon rank-sum test,  $p<0.01$ ,  $n_1=38$ ,  $n_2=155$ ). These data suggest that distrusted trustees use both promise adjustments 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 decision in Game 2 is positively correlated with the amount of \$10 promised in Game 2 ( $p$ -value=0.06), but that message length is not significant. The Game 2 rate of trust-extension for Game 1 distrustful investors was 84.2% (32/38). This is very similar to the original unconditional investment rate of 83.4% in Game 1.

Finally, we evaluate whether promises used by Game 1 distrusted trustees facilitated higher joint payoff in Game 2, and whether those promises were reliable indications of subsequent trustee behavior. Game 2 investments made in previously distrusted trustees paid off for both 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,  $n=32$ ) and the newly trusted trustees earned an average of \$13.12.

On average, newly trusted trustees' promises were veridical – 62.5% (20/32) kept their promises. On the other hand, 37.5% (12/32) of that set broke their promises – more than the 18.8% of trusted trustees who broke their promises in Game 1 (Wilcoxon rank-sum test,  $p=0.05$ ,  $n_1=32$ ,  $n_2=191$ ). One possibility is that the excess promise-breaking of newly trusted trustees reflects a reaction against investors' lack of trust in Game 1, perhaps based on a sense of entitlement for the profits that could have been earned had trust been extended in Game 1. By breaking their promises 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 Game 1 trusted trustees,

than the average earnings of \$10.55 for newly trusted trustees who did not break promises in Game 2. The estimation of specification (2) in Table 5 does not reveal any significant predictor for amount returned by trustees in Game 2.

## 5. Discussions and Conclusions

Opportunities for mutual gains often exist where previous exchange histories have not yet been developed or where trust has been damaged by expectations not met. While promises and apologies appear to be important tools for building and rebuilding trust in these problematic situations, most of the research on these remedial strategies is based on self-report, anecdotal, or archival evidence, or else experiments based on fictional vignettes, videotaped dramatization, or deception. By using a non-deceptive study wherein financially motivated participants used endogenously created and naturally distributed promises and apologies we demonstrate that trustees send cheap signals to encourage new trust and rebuild damaged trust, and that these signals are often effective, leading to benefits for both investor and trustee.

From the egoist perspective of non-cooperative game theory no cooperation is predicted, yet our experiment yielded high rates of trust-extending behavior (e.g. 83.4% in Game 1) and trust-reextension (88% of those who went IN in Game 1 went In in Game 2). There are several non-exclusive accounts for these results. Profit-seeking investors need to trade the risk of trusting under-reciprocators against the risk of not trusting reciprocators.<sup>10</sup> As the efficiency of the investment increases, so do the possible forgone benefits for investors who choose OUT. While the multiplier of 4 used in our study, higher than the multiplier of 3 used in standard trust games, might have contributed to investor willingness to choose IN, these effects are not commonly found across trust games. A meta-analysis by Johnson & Mislin (2011), examining games with different multipliers to evaluate whether a higher multiplier might increase the likelihood of investment, found no effect of multiplier on investors and a strong negative effect on trustworthiness: a higher multiplier *decreases* the amount of money returned by a receiver.

Another possibility, which our data suggests, is that promises in our game may have enhanced the trust extension rate (consistent with Charness & Dufwenberg 2006, who showed IN rates of 74% with promises), and that the apologies issued may have enhanced the trust re-

---

<sup>10</sup> In Game 1, for instance, investors who chose IN received back \$8.19 on average, which is significantly higher than their alternative OUT payoff of \$5. Despite the sizeable rate of under-reciprocation in Game 1, decisions to take on risk by choosing IN yielded higher profits for investors than decisions to not take on risk by choosing OUT.

extension rate, despite broken promises in Game 1. Although we lack experimental controls without signals, within sample comparisons suggest that investors lent credence to specific formulations of trustees' promises, and conditioned their credence to these promises in Game 2 according to cues of trustworthiness and untrustworthiness (i.e., kept promises and broken promise, respectively). We expected that trustees, aware of investor self-interest and motives for critical signal reception, would promise investors mutually beneficial transfers in the range \$6 to \$19. Consistent with results from bargaining games (where even splits are reported as modal offers that also tend to be accepted; see Guth et al. 1982; Guth & Tietz 1986; Carter & Irons 1991; Prasnikar & Roth 1992), even-split promises, which lay close to the middle of the predicted range, elicited more trust-extension than uneven-split promises among dyads with no history of trust-based exchange.

We have argued that as long as the truth value of a signal can be reliably estimated, and updated in tandem with estimates of the signaler's trustworthiness, cheap-to-produce signals such as promises can facilitate coordination and cooperation. In the context of repeated interactions, promises and apologies should be less trusted when issued by trustees whose past promises and apologies were followed by untrustworthy behavior. The proverbial boy who cried wolf illustrates this principle in the domain of predator calls. But does the principle apply in the domain of social exchange? Our experimental design did not include a third game, so we cannot know whether the investors who suffered broken promises in Game 1, were apologized to, and again suffered broken promises in Game 2, would have discounted further apologies. Future research with similar designs but more than two successive games is needed to test the prediction that the credibility attributed to apologies would be recalibrated based on subsequent behavior by offenders. However, our results do provide a partial answer to the question of how signal credibility is calibrated by relevant behavioral cues: as evidenced by Game 2 investment rates, Game 2 promises issued by trustees who previously returned less than promised were given less credence than the Game 2 promises issued by trustworthy trustees.

Nevertheless, among investors whose trust was damaged in Game 1, messages with apologies elicited more re-extension of trust than messages without apologies. While Game 2 promise upgrades might signal intention to provide an economic contribution towards restituting the previously promised amount lost (i.e., *atonement*: a repair done for the sake of a damaged relationship), they could also be attempts at *coercion*: promise-breakers' coaxing efforts –

calibrated to compensate for the investor's expectation that their Game 2 promises would be "exaggerated" (i.e., as they were in Game 1). The former explanation suggests an upgraded regard for the investor, the latter a strategically selfish regard. Evidence from our experiment suggests that most promise breakers upgraded their Game 2 promises out of selfish-regard, since the majority (60%) of promise-breakers who were invested in again went on to break promises a second time.

We suggest that the rate of trust re-extension seen for trustees who turned out to be repeat promise-breakers may have been lower 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 to sanction cheating. We suspect that in the "real world" of non-anonymous and face-to-face interactions, persuasive messages like promises and apologies are likely more effective and less likely to lead to further damaged trust because receivers can evaluate the veracity of verbal messages based on their correspondence with other reliable cues and signals (e.g., past demonstrations of trust or trustworthiness, facial expressions, tone of voice, eye movements, body language).<sup>11</sup>

From this study we see evidence indicating how personal exchanges are often based around establishment of trust via cheap-to-produce verbal signals, and how these 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 signals such as verbal claims about reputation, verbal contracts, and apologies.

---

<sup>11</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.

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 offense and any regret or sorrow stemming from having caused the offense (such as through some form of apology). An optimistic perspective on relationships fraught with damaged trust recognizes that they actually represent opportunities to develop better relations than previously established. Second, to persuade and assure victims that relationship repair is possible the offender should signal (such as with a personalized message) an indication of the value that is recognized in the other, which stems from an internal recalibrations and commits one (such as with a promise) to expectations of future cooperative behavior. In signaling recognition of relationship value it is important not to express a selfish perspective, but instead a shared-welfare or other-regarding perspective. Third, to actually begin the process of changing and redefining the relationship, an offender must be willing to pay costs to expeditiously correct the previous imbalance of welfare (thereby increasing the offended party's welfare), or else to sacrifice wealth or status (thereby decreasing one's own welfare). When corrective actions cannot be immediately 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 general conclusions arrived at by Lazare (2004).

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 practices of firms, religious clergy, and military relations. We strongly encourage further efforts to uncover effective strategies for building trust where previous trust-based exchange histories had not been developed, or where trust had been damaged by reciprocation failure.

## References

- Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science* 242:1385-1389
- Balliet, D. (2010). Communication and cooperation in social dilemmas: a meta-analytic review. *Journal of Conflict Resolution* 54(1): 39-57
- 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
- Benoit, W.L., & Drew, S. (1997). Appropriateness and effectiveness of image repair strategies. *Communication Reports*, 10:153–163.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior* 10:122-142
- 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. & Feltoovich, 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.R., Croson, R. & Johnson, E.J. (2006). Let's Get Personal: An International Examination of the Influence of Communication, Culture, and Social Distance on Other Regarding Preferences. *Journal of Economic Behavior and Organization*, 60(3):373-398.
- Carter, J.R. & Irons, M.D. (1991). Are economists different, and if so, why? *Journal of Economic Perspectives* 5:171-177.
- Charness, G. & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica* 74:1579-1601.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20: 37-46.
- 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.
- 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.
- Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution* 2:256-79.
- Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations* 13:123-3
- Diggle, 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.
- Farrell, J. & Rabin, M. (1996). Cheap Talk. *Journal of Economic Perspectives* 10 (3): 103–118.
- Fischbacher, U. & Utikal, V. (2010). On the acceptance of apologies. *Research Paper Series Thurgau Institute of Economics and Department of Economics* No. 53.
- 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.
- Guth, W.R., Schmittberger, W.R., Schwartz, B. (1982). An experimental analysis of Ultimatum Games. *Journal of Economic Behavior and Organization*. 3:367-388.
- Guth, W.R. & Tietz, P. (1986). Acutioning ultimatum bargaining positions: how to act if rational decisions are unacceptable. In R.W. Scholz (Ed.) *Current Issues in West German Decision Research* (pp. 173-185). Frankfurt: P. Lang.
- 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 the emotions as guarantors of threats and promises. *UCLA Department of Economic Working Paper #337* (August).
- Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *American Economic Review*, 86(3), 653-660.
- Houser, D. & Xiao, E. (2011). Classification of natural language messages using a coordination game. *Experimental Economics* 14:1-14
- Johnson, N.D. & Mislin, A.A. (2011). Trust games: a meta-analysis. in *Journal of Economic Psychology* 32:865-889
- Kelly, R. L. (1995). *The foraging spectrum*. Washington, D.C.: Smithsonian Institution Press.
- 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., 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.
- 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.
- Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: confusion or a heuristic? *Evolution and human behavior*, 21(6), 411-427.
- Knight, R.F. & Pretty, D.J. (1999). Corporate catastrophes, stock returns, and trading volume. *Corporate Reputation Review*, 2(4):363-378
- Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology*. Sage Publications: Thousand Oaks, CA.
- Lachmann, M., Számadó, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences*, 98(23), 13189–13194.
- Landis, J.R. & Koch, G. (1977). An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics*, 33. 363-374.
- Lazare, A. (2004). *On Apology*. New York: Oxford University Press
- 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.
- 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.
- Masclet, D., Noussair, C., Tucker, S., & Villeval, M. (2003). Monetary and non-monetary punishment in the voluntary contributions mechanism. *The American Economic Review* 93(1):366-380.
- Maynard-Smith, J. (1982). Do animals convey information about their intentions? *Journal of Theoretical Biology*, 97:1–5.
- 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.
- 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.
- 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* 41(2):323-331.



- Prasnikar, V. & Roth, E. (1992). Considerations of fairness and strategy: experimental data from sequential games. *Quarterly Journal of Economics* 106: 865-888.
- 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.
- Rohwer, S. (1977). Status Signaling in Harris Sparrows: Some Experiments in Deception. *Behaviour*, 61(1/2), 107-129.
- 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
- 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.
- 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
- Sinaceur, M., & Tiedens, L. Z. (2006). Get mad and get more than even: When and why anger expression is effective in negotiations. *Journal of Experimental Social Psychology*, 42, 314 – 322.
- Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk Analysis*, 13:675-682
- 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. (1993). The fallacy of conventional signalling. *Philosophical Transactions of the Royal Society London, Series B*, 340:227–230.
- 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 &lt; 9</i>	-0.83***	-4.81***	-1.77***	-6.92**
[dummy for promises less than 9 in Game1]	(0.21)	(1.07)	(0.41)	(2.77)
<i>Promise1 &lt; 9 × Promise1</i>		0.60***		0.72*
[amount promised in Game1 for promises < 9]		(0.16)		(0.38)
<i>Promise1 &gt; 11</i>	-1.01**	8.44	-4.52***	12.30*
[dummy for promises greater than 11 in Game1]	(0.41)	(5.19)	(0.95)	(6.62)
<i>Promise1 &gt; 11 × Promise1</i>		-0.55*		-1.11**
[amount promised in Game1 for promises > 11]		(0.30)		(0.43)
<i>Constant</i>	1.35***	1.35***	8.81***	8.81***
	(0.15)	(0.15)	(0.21)	(0.21)
Observations	229	229	191	191

\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%. Standard errors in parenthesis.

**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 &lt; 9</i>	-0.05	-0.07	-0.85	-0.97
[dummy for promises less than 9 in Game1]	(0.36)	(0.36)	(0.79)	(0.78)
<i>Promise1 &gt; 11</i>	-0.37	-0.21	2.67*	1.73
[dummy for promises greater than 11 in Game1]	(0.65)	(0.70)	(1.45)	(1.47)
<i>Return1</i>	0.00	-0.02	0.02	0.03
[return in Game1]	(0.12)	(0.12)	(0.26)	(0.26)
<i>Broken1</i>	-0.80*	-0.84*	0.86	1.86
[broken promise in Game1]	(0.48)	(0.49)	(1.33)	(1.36)
<i>Broken1 × (Promise1-Return1)</i>	-0.07	-0.10	-0.58*	-0.64**
[extent of under-return relative to Game1 promise]	(0.11)	(0.12)	(0.30)	(0.30)
<i>Distrusted1</i>	-0.22	-0.32	-1.53	-1.35
[not trusted in Game1]	(1.06)	(1.10)	(2.39)	(2.38)
<i>Promise2 &lt; 9</i>	-1.37***	-2.13**	-1.66**	-1.47
[dummy for promises less than 9 in Game2]	(0.32)	(0.88)	(0.74)	(2.79)
<i>Promise2 &lt; 9 × Promise2</i>		0.11		-0.01
[amount promised in Game2 for promises < 9]		(0.12)		(0.38)
<i>Promise2 &gt; 11</i>	-0.14	2.20	0.78	-12.01**
[dummy for promises greater than 11 in Game2]	(0.42)	(1.90)	(1.05)	(4.64)
<i>Promise2 &gt; 11 × Promise2</i>		-0.14		0.80***
[amount promised in Game1 for promises > 11]		(0.11)		(0.28)
<i>Constant</i>	2.05*	2.26*	8.98***	8.83***
	(1.13)	(1.17)	(2.51)	(2.48)
Observations	229	229	200	200

\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%. Standard errors in parenthesis.

**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>	0.15	0.11	-0.01	-0.05
[return in Game 1]	(0.15)	(0.15)	(0.21)	(0.21)
<i>Message</i>	0.63*	0.61	-0.24	-0.24
[message with content]	(0.38)	(0.38)	(0.63)	(0.63)
<i>Promise2&lt;9</i>	-1.12***	-4.39**	-2.23***	-9.13
[dummy for promises less than 9 in Game2]	(0.43)	(2.21)	(0.69)	(5.57)
<i>Promise2&lt;9 × Promise2</i>		0.45		0.92
[amount promised in Game2 for promises < 9]		(0.30)		(0.74)
<i>Promise2&gt;11</i>	-1.83**	-1.82**	9.12***	9.08***
[dummy for promises greater than 11 in Game2]	(0.80)	(0.81)	(2.01)	(2.00)
<i>Constant</i>	0.29	0.63	9.23***	9.58***
	(1.42)	(1.47)	(2.04)	(2.06)
Observations	155	155	143	143

\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%. Standard errors in parenthesis.

**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>Return1</i>	0.10	-0.11	0.69***	1.15**
[return in Game 1]	(0.08)	(0.15)	(0.24)	(0.46)
<i>Promise2-Promise1</i>	0.21*		-0.46	
[promise adjustment]	(0.13)		(0.45)	
<i>Promise1-Return1</i>		-0.21*		0.46
[extent of under-return relative to Game1 promise]		(0.13)		(0.45)
<i>Apology</i>	1.88***	1.88***	2.69	2.69
[message with apology]	(0.68)	(0.68)	(3.19)	(3.19)
<i>Promise2</i>	-0.09	0.13	0.37	-0.09
[amount promised in Game 2]	(0.12)	(0.08)	(0.48)	(0.29)
<i>Constant</i>	-0.56	-0.56	-1.90	-1.90
	(1.58)	(1.58)	(6.24)	(6.24)
Observations	36	36	25	25

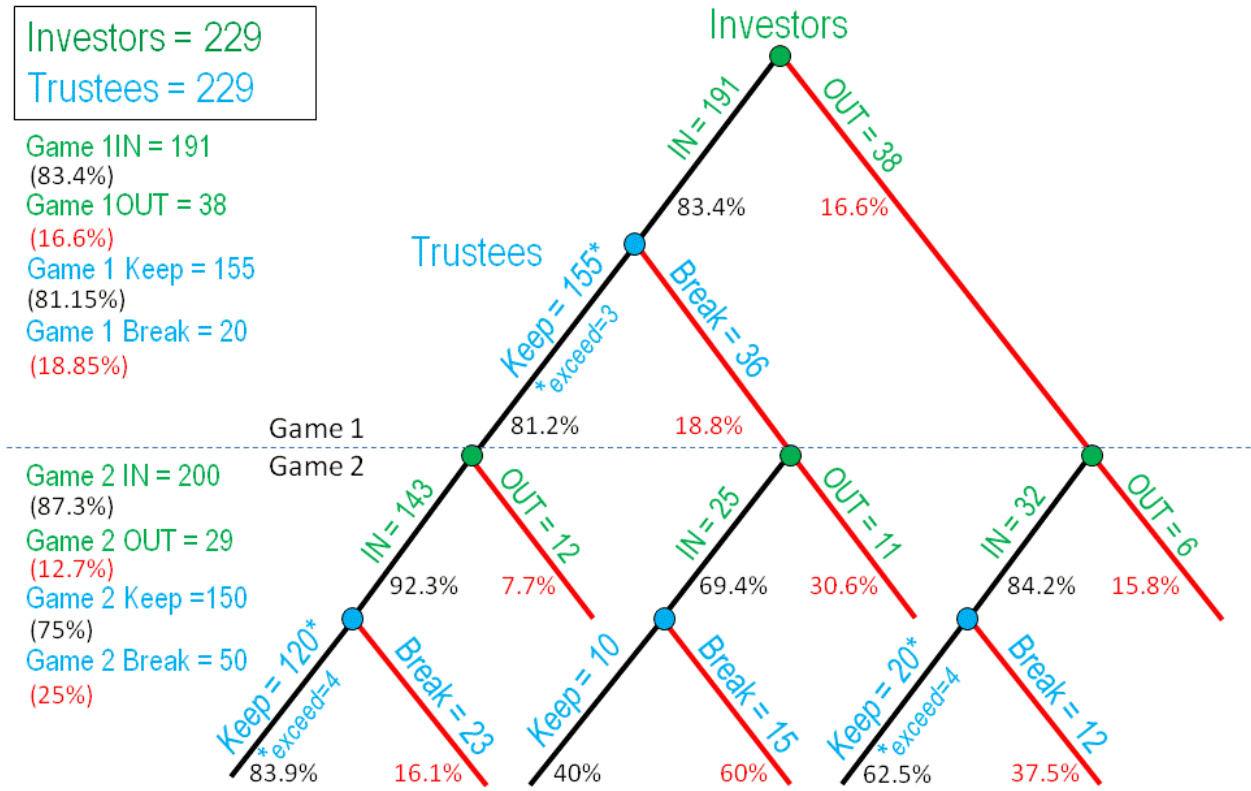
\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%. Standard errors in parenthesis.

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

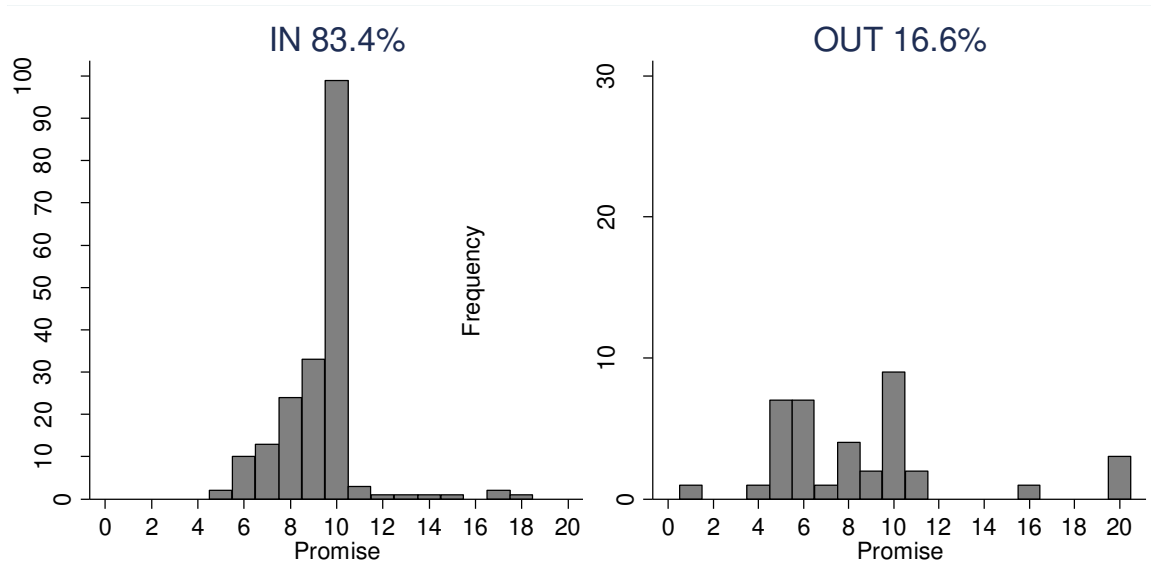
Dependent variable	<i>Invest2</i>	<i>Return2</i>
Specification	(1)	(2)
	Probit	OLS
<i>Promise1=10</i>	-0.86	2.51
[dummy for promises equal to 10 in Game1]	(0.70)	(2.16)
<i>Wordcount</i>	0.02	0.00
[number of words]	(0.01)	(0.04)
<i>Promise2=10</i>	1.40*	0.08
[dummy for promises equal to 10 in Game2]	(0.76)	(1.89)
<i>Constant</i>	0.38	6.28***
	(0.47)	(1.73)
Observations	38	32

\* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%. Standard errors in parenthesis.

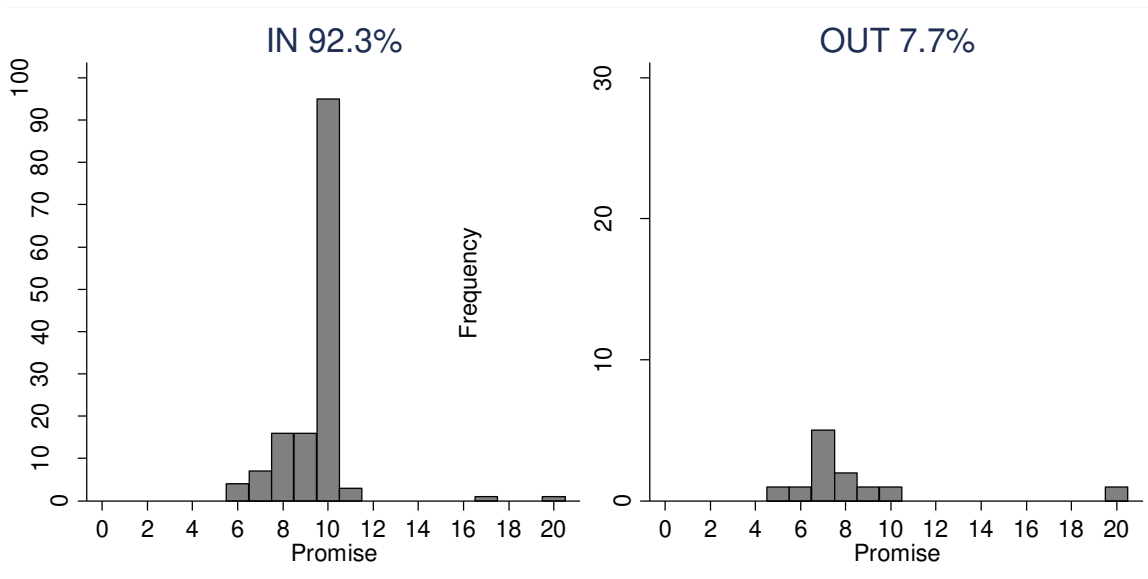
**Figure 1: Aggregate Distribution of Decisions in Games 1 and 2**



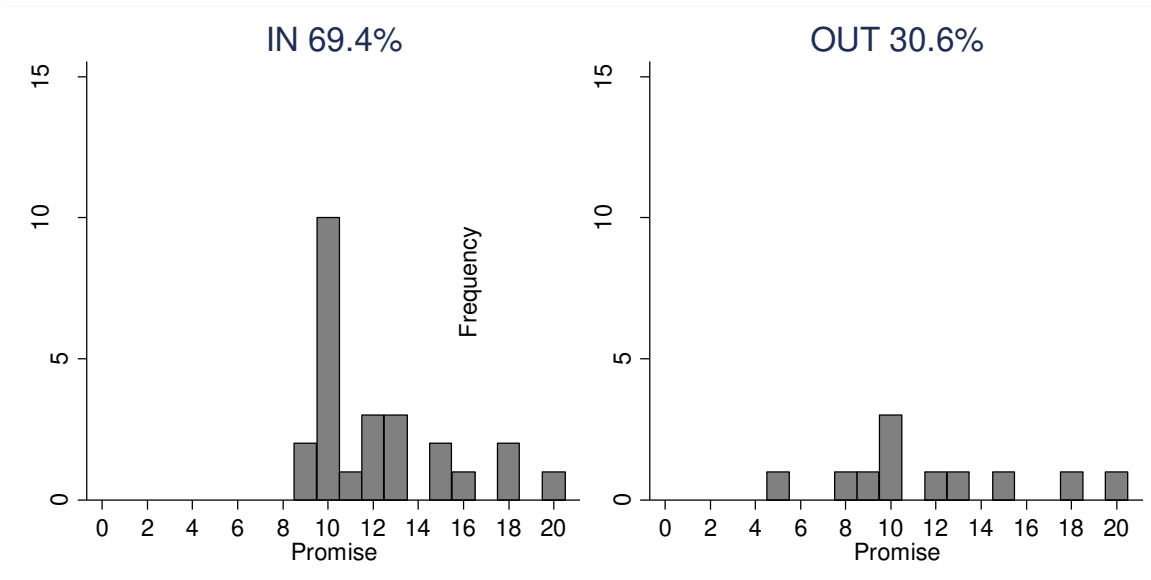
**Figure 2: Distribution of Promises in Game 1 (Resulting in IN or OUT)**



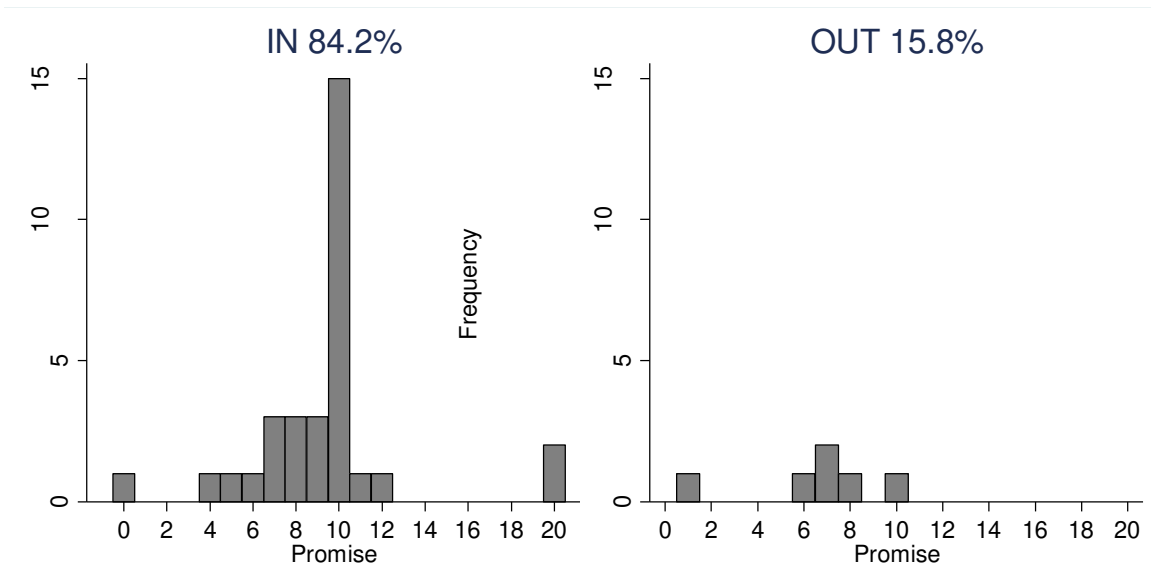
**Figure 3: Distribution of Promises in Game 2 (Resulting in IN or OUT) by Game 1 Promise-Keepers**



**Figure 4: Distribution of Promises in Game 2 (Resulting in IN or OUT) by Game 1 Promise-Breakers**



**Figure 5: Distribution of Promises in Game 2 (Resulting in IN or OUT) by Game 1 Distrusted Trustees**



## Appendix A: Instructions for Trust Game

### 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.

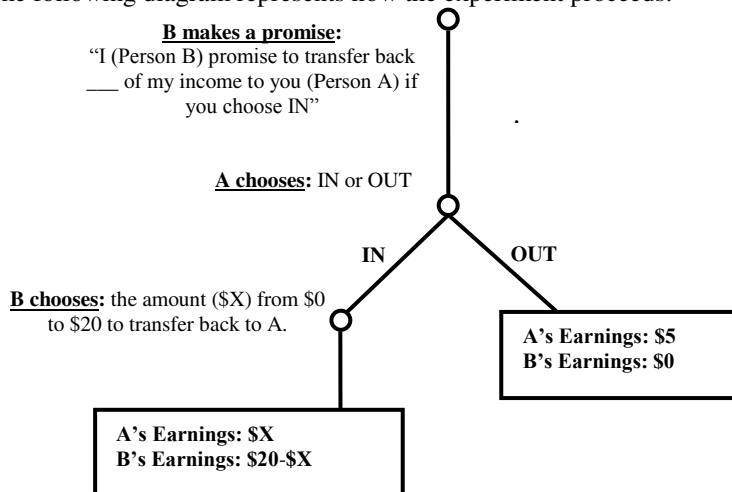
Having received a statement from B, A indicates whether he or she chooses IN or OUT. If A chooses OUT, A receives \$5 and B receives \$0. If A chooses IN, then B receives \$20 income. In such a case, after receiving \$20 income, B must choose a dollar amount from \$0 to \$20 to transfer back to A.

### 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: Instructions for Coders

### INSTRUCTIONS

You will be playing a MATCHING GAME with two other students, called your partners. Your payment for the matching game depends not only on what you do, but also on what your partners do, and on the rules of the matching game described below. Make sure you understand the rules. Feel free to ask us questions as they arise, by raising your hand. Please do not speak to other students during the experiment. You will receive \$7 for participating in this session. You may also receive additional money, depending on the decisions you and the two other students make. Upon completion of the session, this additional amount will be paid to you individually and privately.

### BACKGROUND

Over the last two years, many Chapman students like you came to our lab to participate in economic experiments which we call “trust games.” In these games, each participant was paired with another participant. All participants interacted via computer terminals and the interaction was anonymous. That is, no participant knew the identity of the person with whom they were paired. In each pair, one person had the role of A, and the other had the role of B.

The trust game proceeded in several stages. First, by choosing a dollar amount from \$0 to \$20, B indicated the proportion of a possible \$20 income that he or she promised to transfer back to A, should A choose IN. Specifically, B completed the following statement: “I (Participant B) promise to transfer back \_\_\_ of my income to you (Participant A) if you choose IN”. Both B and A understood these promises to be non-binding. The computer conveyed B’s statement to A. Having received a statement from B, A chose either IN or OUT. If A chose OUT, A received \$5 and B received \$0. If A chose IN, B received \$20 income. In such a case, after receiving \$20 income, B chose a dollar amount from \$0 to \$20 to transfer back to A.

Then we presented exactly the same trust game to the same participants once again. Neither participant A nor participant B knew that there would be an unexpected second trust game. After learning the outcome of the first trust game (from now on we will call it Game 1) and after the unexpected second game (Game 2) was announced, but before it commenced, B could send a one-way message to A – saying anything that did not personally identify them.

You are NOT playing these trust games! Instead, you and your partners will play the matching game.

### MATCHING GAME

In the matching game you will be examining the messages sent by B to A. Specifically, we will present to you 36 messages that were sent by B to A after B learned that there would be a Game 2. In all 36 instances, in Game 1, B promised to return a certain amount and then after A chose IN, B chose to transfer back to A an amount that was lower than the amount promised. We will call these participants “promise-breakers,” since they broke their promises in Game 1 by returning less than they promised they would.

Your job, and that of your partners in the matching game, is to make judgments about whether these written messages sent by promise-breakers after Game 1 and before Game 2 qualify as either of two definitions of “apologies.” You and your partners are trying to match your judgments about each of the messages. The more messages you match, the more you earn. Another way of thinking about this is as follows: You are trying to guess how your two partners will judge each message you all see and to make matching judgments yourself. The more accurate your guess about their judgments, the more money you can make.

We will provide you with participant-specific data on the Game 1 promise made, the Game 1 return made, and the subsequent message sent. Then, you and your partners will have to answer YES or NO to the following two questions:

Does the message qualify as an apology according to apology definition 1?

Does the message qualify as an apology according to apology definition 2?

We cannot tell you how to play this matching game. Obviously, a lot of doing well depends on what your partners do. However, we recommend that, as you read each message, you and your partners pay attention to what

you expect most people would interpret according to definition 1 and definition 2. We recommend that you DO NOT think too deeply about it. If you and your partners coordinate on the same intuitive approach, your initial interpretations should be most likely to pay off in this game. They are more likely to be similar for you and your partners than alternative interpretations produced by any deep thinking.

## TWO DEFINITIONS OF APOLOGY THAT YOU WILL MAKE QUALIFYING JUDGEMENTS ABOUT

*Apology definition 1:* an explicit or implicit acknowledgment of offense which another has received

*Apology definition 2:* an explicit or implicit acknowledgment of offense which another has received, along with remorse, regret, or sorrow stemming from acknowledgment of offense

## EARNINGS

You earn money in the matching game by making a judgment that agrees with the majority judgment of the group. In other words, if you match your judgment with at least one other partner – the two of you who have matched have formed a majority (i.e., at least 2 out of 3) and will be rewarded with \$5 if the item you matched on is chosen for payment. If your judgment about a message does not match either of your partners', you will not earn a money reward for that judgment.

Specifically, for each of the 36 messages you view, you will make YES or NO judgments in each of two columns that follow. In the first column you will circle either YES or NO, depending whether you think that the message content qualifies as an apology according to definition 1. In the second column you will circle either YES or NO, depending on whether you think that the message content qualifies as an apology according to definition 2. Then for each individual message we will compare your judgment with judgments of the two other students. To calculate your earnings at the end of the experiment, we will randomly choose 3 out of 36 judgments in the first column and 3 out of 36 judgments in the second column for payment, using a bingo cage which contains balls numbered from 1 to 36. Since for each correct judgment you can earn \$5, you can earn up to \$30 for the matching game in addition to the \$7 participation fee.

## SUMMARY

We will distribute to you a set of 36 Game 1 promise-breaker's promise and return decisions and subsequent messages. In all, you will view 36 messages. Each message will be followed by two columns, each with YES or NO. In the first column you will circle either YES or NO, depending whether you think that the message content qualifies as an apology according to definition 1. In the second column you will circle either YES or NO, depending on whether you think that the message content qualifies as an apology according to definition 2.

Since we will choose three judgments in column 1 and three judgments in column 2, each potentially earning a reward of \$5.00, you can earn up to \$30 for the matching game. You additionally will earn \$7 for showing up on time and participating.

## Appendix C: Promises and Messages

**Table C1: 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.	6	YES	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.	43	YES	YES	15	YES	3
10	0	Hey im sorry about that I didn't realize there was going to be another round.! Let me make things right.	20	YES	YES	15	YES	5
10	9	i'll do the same deal as last time, sound fair?	10	NO	NO	15	YES	9
10	7	to even out i will give you 13 and i will take 7	13	YES	YES	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 ;)	43	YES	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.	27	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!	88	YES	NO	12	YES	12
8	5	I only sent less than promised because I wanted to see what would happen	14	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	19	YES	NO	9	YES	9
13	9	lets split the money 10 and 10	7	YES	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.	47	YES	YES	9	YES	9
10	7	I didn't know we were repeating this. This time I really will split 50/50 ;)	15	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.	24	YES	YES	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.	35	YES	YES	18	YES	18
11	10	10/10?	1	YES	NO	12	YES	7
8	0	sorry about last time i feel bad.....50/50 this time?	10	YES	YES	10	YES	0
8	7	Strategy ;)	1	YES	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 possible	40	YES	YES	12	YES	0
8	6	My sincerest appologies on that last one...I do feel quite guilty ¶ and I assure you that this	34	YES	YES	16	YES	1

		time I shall keep my promise with utmost integrity. You have my word as an honest gentleman.						
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.	24	YES	YES	13	YES	13
5	1	my bad...	2	YES	NO	10	YES	8
10	3	I'll transfer back more money this time ¶ actually \$10	9	YES	YES	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.	28	YES	NO	10	YES	10
10	0		0	NO	NO	20	YES	0
10	0	This time I'll give you what I promise. Sorry!	9	YES	YES	10	NO	
10	0	ok for real this time haha. The first time was a joke lol	13	YES	NO	15	NO	
17	0	May God bless you	4	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.	34	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.	41	YES	YES	18	NO	
10	5	Even though I was dectieful ¶ you were no worse off then had you picked OUT. The other option would have still led you to \$5.	25	YES	NO	5	NO	
12	5		0	NO	NO	20	NO	
18	4		0	NO	NO	12	NO	
7	6		0	NO	NO	9	NO	
10	5		0	NO	NO	8	NO	
10	7		0	NO	NO	10	NO	

**Table C2: 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 accept ¶ 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 receive 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 C3: 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 end.		9	YES	9

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	