9-2006

Promises and Lies: Restoring Violated Trust

Maurice E. Schweitzer
University of Pennsylvania

John C. Hershey
University of Pennsylvania

Eric T. Bradlow
University of Pennsylvania

Recommended Citation

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and Lies: Restoring Violated Trust. *Organizational Behavior and Human Decision Processes, 101* (1), 1-19. http://dx.doi.org/10.1016/j.obhdp.2006.05.005

Promises and Lies: Restoring Violated Trust

Abstract

Trust is critical for organizations, effective management, and efficient negotiations, yet trust violations are common. Prior work has often assumed trust to be fragile—easily broken and difficult to repair. We investigate this proposition in a laboratory study and find that trust harmed by untrustworthy behavior can be effectively restored when individuals observe a consistent series of trustworthy actions. Trust harmed by the same untrustworthy actions *and deception*, however, never fully recovers—even when deceived participants receive a promise, an apology, and observe a consistent series of trustworthy actions. We also find that a promise to change behaviour can significantly speed the trust recovery process, but prior deception harms the effectiveness of a promise in accelerating trust recovery.

Keywords

Deception, negotiation, trust, trust repair

Disciplines

Experimental Analysis of Behavior | Social and Behavioral Sciences

Promises and Lies: Restoring Violated Trust

Maurice E. Schweitzer
566 JMHH, OPIM
3730 Walnut Street
The Wharton School, University of Pennsylvania
Philadelphia, PA 19104-6340
Phone/Fax: 215.898.4776/3664
E-mail: Schweitzer@wharton.upenn.edu

John C. Hershey
568 JMHH, OPIM
The Wharton School, University of Pennsylvania
Philadelphia, PA 19104
Phone/Fax: 215.898.5041/3664
E-mail: Hershey@wharton.upenn.edu

Eric T. Bradlow
761 JMHH, Marketing and Statistics
Wharton School, University of Pennsylvania
Philadelphia, PA 19104
Phone/Fax: 215.898.8255/2534
E-mail: EBradlow@wharton.upenn.edu

Acknowledgement: We thank the Zicklin Center for Business Ethics Research for support. We thank Samantha Rudolph and Clifford Lou for research assistance.

Abstract

Trust is a critical catalyst for market transactions and effective management, yet trust violations in both economic markets and organizations are common. In this paper, we examine how trust recovers following untrustworthy behavior and deception. Prior work has often assumed trust to be fragile—easily broken and difficult to repair. We investigate this proposition in a laboratory study (n=184) using a modified and repeated version of a trust game (Berg, Dickhaut, & McCabe 1995). We find that trust harmed by untrustworthy behavior can be effectively restored. Trust harmed by the same untrustworthy actions *and deception*, however, never fully recovers—even when deceived participants receive a promise, an apology, and a series of trustworthy actions. Overall, we find that a "cheap talk" promise can speed initial trust recovery, but such a promise followed by a series of observed, trustworthy actions is no more effective in restoring long-term trust than a series of trustworthy actions alone.

Keywords: Deception, Trust Recovery, Trust Game, Cheap Talk

Introduction

Trust is essential for economic and social systems (Donaldson 2001, Knack & Keefer 1997). Trust facilitates market transactions (Bromily & Cummings 1995, Hirsch 1978, Ring & Van de Ven 1992) and enables managers to lead more effectively (Atwater 1988, Bazerman 1994) and negotiate more efficiently (Butler 1999: Schurr & Ozanne 1985). At the same time, however, we know that trust is often violated. Fraud is a common problem in business settings (Santoro & Paine 1993, Business Week 1992, Los Angeles Times 1998) and deception pervades managerial life (Carr 1968, O'Connor & Carnevale 1997, Santoro & Paine 1993, Schweitzer & Croson 1999).

Surprisingly, however, little is known about the consequences of violating trust. While common wisdom presumes that trust violations cause severe relationship damage (e.g., Slovic, 1993), prior work has not examined how trust actually changes over time as a function of different types of violations and attempts to restore it. In this article, we report results from a laboratory study that investigates changes in trust over time. We examine how deception and untrustworthy behavior harm trust and how a promise, an apology, and trustworthy actions repair trust.

Trust

A substantial literature has identified a number of individual and contextual factors that influence trust (Cook & Wall, 1980; Dasgupta, 1988; Deutsch, 1960; Lewicki & Weithoff, 2000; Mayer, Davis, & Schoorman 1995; Ross & LaCroix 1996; Williams, 2001). Most prior research, however, has examined trust as a static construct, and has devoted surprisingly little attention to how trust might change over time, particularly after it has been harmed. There are a few exceptions. Lewicki and Bunker (1996) and Lewicki and Wiethoff (2000) develop theoretical models that consider the implications of trust violations. Their work suggests that trust violations may irrevocably harm trust. They consider different types of trust relationships ranging from emerging or early-stage relationships (e.g. a relationship among strangers) to mature or late-stage relationships (e.g. a relationship among spouses). They postulate that trust violations harm late-stage relationships more than they harm early-stage relationships. Even this work, however, offers little theoretical insight into the nature and limits of trust recovery.

Recent experimental work has significantly expanded our understanding of the mechanics of trust. Some of this work has also begun to examine how trust changes from one period to the next. In several cases, results from experimental work have contradicted important assumptions developed in theoretical models (e.g. Glaeser, Laibson, Scheinkman, & Soutter, 2000; Pillutla, Malhotra, & Murnighan, 2002). For example, while prior theoretical work has identified

familiarity and future interdependence as key factors for trust development (Williams 2001), Ho and Weigelt (2002) demonstrate that trust can develop even among anonymous strangers who have no opportunity for future interactions.

In a second set of experiments, Pillutla, Malhotra, and Murnighan (2002) challenge a common recommendation suggesting that decision makers move in small steps to build trust (Thompson 2001). In their experiments, Pillutla, Malhotra, and Murnighan (2002) measured trustworthy reactions to different types of trust decisions. They find that small trusting actions may actually fail to build a cycle of trusting and trustworthy behavior.

One factor that is likely to play an important role in trust development is the attributions people make for prior behavior (Larrick & Blount, 2002; Pillutla, Malhotra, & Murnighan, 2002). Malhotra and Murnighan (2002) investigate the influence of attributions in trust decisions in a novel experiment. In their experiment they allowed participants to use binding contracts. They find that the use of binding contracts leads to situational, rather than personal, attributions of trustworthiness and that these attributions inhibit trust development. After using a contract, people actually trust their counterpart less.

Related work has examined the effects of communication in a repeated prisoner's dilemma game. This work has found that non-task communication can significantly increase cooperation (see Sally, 1995 for a review). For example, Gibson, Bottom, and Murnighan (1999) and Bottom, Daniels, Gibson and Murnighan (2003) found that apologies help restore cooperation following uncooperative actions.

Our Investigation

In this paper, we focus on trust in emerging relationships (e.g. trust between strangers) and we use experimental methods to investigate changes in trust over time. Our work differs from prior investigations in several important ways. Most importantly, our work is the first to describe the influence of deception on trust over time. We measure both short-term and long-term effects of deception, and disentangle the harmful effects of untrustworthy behavior, deception, and the interaction between prior deception and subsequent promises on trust recovery.

We define trust as the "willingness to accept vulnerability based upon positive expectations about another's behavior" (Rousseau, Sitkin, Burt, & Camerer, 1998). This definition represents a multidisciplinary approach to defining trust (see Hosmer, 1995 and Mayer, Davis, & Schoorman, 1995 for reviews), and in our experiment we measure trust using both behavioral and attitudinal measures.

Our primary measure of trust is passing behavior in a repeated trust game (Berg, Dickhaut, & McCabe 1995). We depict the actual version of the game in Figure 1. In our experiment, participants are told that they will play several rounds of the same game with the same partner. Odd players (our participants) are endowed with $6 in each round and can either "Take $3," "Take $6," or "Pass $6." If the odd player takes money the round ends, and the odd player keeps the amount that s/he took. If the odd players passes $6, the amount of money triples (to $18) and the even player decides how much money to return to the odd player. If odd players have favorable expectations over the amount even players will return, they will be more likely to accept vulnerability (e.g. the chance of having no money returned) and pass $6. Note that the option to take $3 is strictly dominated by the decision to take $6, but affords participants who do not trust their counterpart an opportunity to make an altruistic choice.

All of our participants make decisions as odd players, and receive feedback and prepared messages from even player confederates. We use the strategy method to conduct the experiment so that every participant is exposed to a consistent set of even player actions regardless of their passing or taking decisions. All participants learn, in a round-by-round sequential manner as the game unfolds, that their counterpart chooses untrustworthy actions in the first two rounds (the even player returns $0), and trustworthy actions thereafter (the even player returns $9). We develop our hypotheses with respect to this framework.

Hypotheses

Our hypotheses focus on the effects of communication on the trust recovery process. By examining trust in a repeated game, we measure changes in trust over time.

A strong null hypothesis predicts no effect for any communication. In our experiment, odd player participants receive identical information about their even player counterpart's actual choices (i.e., return $0 in the first two rounds, and return $9 thereafter). All of the communication in our experiment constitutes "mere" or "cheap" talk. It is costless for participants to send messages, and the communication does not allow participants to formalize agreements (see Farrell & Rabin, 1996).

This null hypothesis regarding communication serves as a foil for our main hypotheses. Prior work has, in fact, found that "cheap talk" affects strategic actions in games (Buchan, Johnson, & Croson, 2002; Dawes & Orbell, 1995; Gneezi, 2002; Sally, 1995). This work has focused on the role of "cheap talk" in facilitating cooperation or trust. In this paper, we focus particular attention on the role that "cheap talk" can play in harming and restoring trust.

Deception and Trust Recovery

We first consider the influence of deception on the trust recovery process. In our experiment, half of the participants receive deceptive messages prior to rounds 1 and 2, in which confederate even players indicate that they will return a substantial amount of money in the upcoming round. In fact, all confederate even players return $0 in both round 1 and round 2. That is, in our experiment trust is harmed by both untrustworthy behavior, to which every participant is exposed, and deception, to which only half of the participants are exposed.

In our experiment, both those deceived and those not deceived may or may not receive subsequent communication. This subsequent communication may interact with prior deception in influencing trust over time. We develop hypotheses for these interactions, but for our initial hypothesis concerning deception, we consider only those who receive no further communication beyond round 2.

We expect the combined effects of deception and untrustworthy behavior to harm trust more than untrustworthy behavior alone. The use of deception conveys information about a counterpart's motivation, and while we expect the harmful effects of deception to diminish over time, we expect deception to harm long-term trust.

Hypothesis 1: In the absence of other communication to restore trust, deception harms the long-term level of trust.

Trust Restoring Communication and Trust Recovery

Next, we examine the effects of trust-restoring communication on trust recovery. We consider effects for a promise and for an apology.

A Promise. Prior work suggests that damaged trust may be very difficult to regain. In general, the trust recovery process is assumed to be slow and incomplete (Slovic 1993, Lewicki & Bunker 1996). In this study, we consider the role of trustworthy actions and a promise in facilitating trust recovery. We posit that a promise will facilitate the trust recovery process by clarifying a counterpart's future trustworthy intentions. This proposition is related to a result identified by Ho and Weigelt (2002) in which they found people to be more trustworthy when they were sure about the intentions of their counterpart.

In our study, some subjects received a written promise of cooperation after round 2 and just prior to round 3 (after the two initial rounds of untrustworthy actions). We expect such a promise to increase both initial trust recovery and long-term trust recovery. We examine the influence of a promise on trust recovery in the absence of other communication (e.g., deception).

Hypothesis 2a: In the absence of other communication, a promise to change behavior will facilitate initial trust recovery.

Hypothesis 2b: In the absence of other communication, a promise to change behavior will increase the long-term level of trust.

An Apology. We expect an apology, coupled with a promise, to restore trust more quickly and more completely than a promise alone. That is, we expect people to respond to a written apology by becoming more trusting in both the short- and long-term. This hypothesis is consistent with prior work that has found that apologies mitigate negative reactions (Ohbuchi, Kameda, Agarie 1989, Shapiro 1991) and help to reestablish cooperation (Gibson, Bottom, & Murnighan 1999). In our experiment, some participants receive an apology coupled with a promise to cooperate, prior to round 3. (Recall that other participants receive either no communication or a promise alone.) To test our hypotheses regarding an apology, we compare participants who received a promise and an apology to those who received only a promise with no other communication.

Hypothesis 3a: In the absence of other communication, an apology and promise will repair initial trust more than a promise alone.

Hypothesis 3b: In the absence of other communication, an apology and promise will increase the long-term level of trust more than a promise alone.

Interaction between Deception and Trust Restoring Communication

We consider the interaction between deception and trust restoring communication—a promise alone or a promise and an apology. Although a purely rational economic agent would discount all communication in this experiment, we expect prior deception to significantly harm the effectiveness of subsequent trust restoring communication.

A Promise and Deception: We consider the interaction between a promise and prior deception. In the short term, we expect promises to restore trust more following no deception, because a promise following deception is likely to be significantly discounted. The long-term effects of a promise following deception relative to a promise following no deception will depend on the extent to which trustworthy actions restore credibility in the promise and restore overall trust. While a promise may play a significant role in restoring trust for those who are deceived, it still may not be as effective as for those who were never deceived.

Hypothesis 4a: Prior deception will harm the initial effectiveness of a promise in restoring trust.

Hypothesis 4b: Prior deception will harm the long-term effectiveness of a promise in restoring trust.

An Apology and Deception: The final set of hypotheses considers the interaction between an apology and deception. These hypotheses investigate whether an apology coupled with a promise restore trust more or less, relative to a promise alone, following deception than following no deception. An apology implies atonement for past communication as well as past behavior, and thus an apology may restore trust more effectively following deception than following no deception, relative to a promise alone.

Hypothesis 5a: The initial effect on trust of an apology coupled with a promise, relative to a promise alone, will be greater following deception than following no deception.

Hypothesis 5b: The long-term effect on trust of an apology coupled with a promise, relative to a promise alone, will be greater following deception than following no deception.

Experimental Methods

We conducted an experiment to examine trust recovery. Participants in our study made a series of trust decisions in the game depicted in Figure 1. An important feature in our experiment is that every participant plays the role of the odd player. We manipulated even player actions and use the strategy method to provide participants consistent feedback; the even player chooses to return \$0 the first two rounds and return \$9 for rounds three through seven regardless of the actions taken by the odd player. That is, every participant observes the same set of even player actions even if they decide not to pass. This aspect of our design is critical to keeping the information each participant receives constant. In our discussion section, we consider some implications of this design with respect to our use of deception and the presence of feedback that facilitates trust recovery.

The experiment includes three separate phases. In the first phase, involving the first two rounds (r = 1 to 2), all participants observed untrustworthy actions. In the second phase, involving the middle four rounds (r = 3 through 6), all participants observed trustworthy actions. In this experiment, we added a third phase, the final round (r = 7), to account for a potential end-game effect.

Recruiting Participants. We recruited participants for a 1½-hour experiment using class announcements. Participants were told that they would have the opportunity to earn money and

that the amount they earn would depend upon their own decisions, the decisions of others, and chance.

Assignment of Conditions. Upon arrival to the experiment, participants were randomly separated into two different rooms. Within each room, participants were then assigned randomly to a treatment condition and a pairing number.

Trust Game. Participants were told that they would play several rounds of the game depicted in Figure 1, that one of these rounds would be randomly selected using a draw from a bingo cage, and that they would be paid the amount they earned for that round.

Prior to the game, participants were given Figure 1 as well as an explanation of the game. Following the explanation, participants answered six comprehension questions. The comprehension questions were designed to accomplish two aims: first, to ensure that participants understood the game; second, to give participants the assurance that their counterpart understood the game. An experimenter individually checked participants' answers and explained the game again to anyone making a mistake. Mistakes were very rare.

Design. Participants were told that they would play the same game with the same partner for several rounds. They were not told the total number of rounds they would play, but they were told that there would be at least seven rounds, and that both odd and even players would receive an announcement indicating the last round just prior to that round. We use this approach to disentangle end-game behavior from the main part of the experiment.

The even player actions that the odd players (our participants) observe are held constant across conditions. Groups of odd player participants were randomly assigned to one of six between-subject communication conditions, which we depict in Figure 2. These conditions result from a 2x3 design: two deception conditions (Deceptive messages prior to rounds 1 and 2, No messages prior to rounds 1 and 2) and three apology conditions (No message prior to round 3, Promise alone prior to round 3, Promise and apology prior to round 3). In every condition, odd players received a message sheet prior to making their trust game decision in round 1, round 2, and round 3. Participants were informed that communication was not allowed after round 3. The top portion of each message sheet asks the even player whether or not they want to send a message. In the no message conditions the "no" box was checked and no message was included on the sheet. In the other communication conditions the "yes" box was checked and a handwritten text message was included at the bottom of the sheet.

The two deception conditions dictated communication prior to rounds 1 and 2. In the deception condition, the odd player received two false statements. The round 1 message read, "If you pass to me I'll return $12 to you." The round 2 message read, "Let's cooperate. I'll really return $12 this time." In the second deception condition, the no deception condition, the odd player received a message sheet prior to round 1 and round 2 indicating that the even player chose not to communicate.

Three apology conditions dictated communication prior to round 3. In the promise and apology condition the message read, "I really screwed up. I shouldn't have done that. I'm very sorry I tried taking so much these last 2 rounds. I give you my word. I will always return $9 every round, including the last one." In the promise alone condition the message read, "I give you my word. I will always return $9 every round, including the last one." In the no promise condition the even player chose not to communicate prior to round 3.

Procedure

Participants made several rounds of trust game decisions. After each round participants completed a brief post-decision survey. This survey asked participants a set of questions including how much they trust their partner. After participants completed the post-decision survey, and had waited an additional 2 to 3 minutes, they received feedback regarding their counterpart's choice for that round (the amount their counterpart returned or would have returned if they, the odd player, had passed).

Prior to making a decision in round 7, we announced, "This will be the last round. Both odd and even players receive this same announcement." Participants then made their final trust game decision and completed their seventh post-decision survey. They waited two to three minutes, received feedback regarding their counterpart's choice for the final round ("Return $9"), and then completed a final survey. The final survey asked them how much they trusted their partner, what they thought their partner was trying to do during the game, and demographic questions.

Payment. After participants completed the final survey, we randomly selected one of the seven rounds using a draw from a bingo cage and paid participants based upon the amount of money they earned for that round. To mitigate participants' potential feelings of disappointment for not having been paired with a real partner, we announced an unanticipated $5 show-up fee that we added to their total payment.

Dependent Measures. We measure trust in two ways. First, we measure trust behavior as the binary decision to pass or take in each of the seven rounds. Second, we use the seven post-decision survey responses that asked participants how much they trust their counterpart. By measuring trust in these two ways, we observe the trust recovery process in actual passing decisions, stated trust intentions, and a comparison of the two.

Investigating the correspondence between passing decisions and self-reported trust ratings is important because prior work has found that decisions, such as the passing decision we model in this study, are influenced by a number of social preferences including preferences for social welfare, recipriocity, and altruism (Andreoni & Vesterlund, 2001; Ashraf, Bohnet & Piankov, 2003; Cox, 2003; Charness & Rabin, 2002). In our work, however, we find an extremely close link between trust ratings and passing decisions. We describe this relationship in the results section.

We use a parametric approach to model our key dependent variable, to pass or not to pass, as a binary decision. We use a parametric approach for two main reasons. First, our parametric approach enables us to fit meaningful variables, such as the long-run asymptote of trust recovery, that are not identified by using standard econometric models. Second, our parametric approach enables us to fit a relatively parsimonious model. In contrast to the model we fit, a traditional parametric model would require 42 parameters to model passing decisions for each of the six conditions across the seven rounds.

In our model, we define P_{irc} as the probability that person i trusts ("passes") in round r following communication condition (e.g. a promise and an apology) c, c = 1 to 6; note, however, that we can (and do) directly adapt this model for a Likert rating dependent variable (e.g. how trusting someone is).

Model for the Experiment. The model we fit is the following:

$$P_{irc} = \text{logit}^{-1}\left(a_i + \begin{cases} X_c(r-3) + (A_c - B_c) & \text{for } r = 2 \\ A_c - B_c \exp\{-d_c(r-3)\} & \text{for } r = 3 \text{ to } 6 \\ A_c - B_c \exp\{-3d_c\} + Y_c & \text{for } r = 7 \end{cases}\right) \quad (1)$$

We use a logit transformation to map our model values onto the [0,1] probability scale to represent the probability of passing. Prior work has identified individual variation in predispositions to trust other people (Rotter 1971), and thus, in our model, we include an individual-level intercept parameter a_i.

The first two piecewise components of our model correspond to the two places where trust recovery might take place. Communication alone (e.g. an apology) may repair trust (at the beginning of round three), as may subsequent trustworthy behavior. We depict these periods and the corresponding pieces of the model in Figure 3. The third piecewise component of the model corresponds to the end game (round seven).

In this model, X_c represents the change in trust behavior due to communication alone prior to round three. The parameter A_c represents the long-run asymptote of trust recovery (i.e. P_{irc} as r ? 8), and the parameter B_c represents the amount of long-term trust recovery due to trustworthy action. Note that the difference $(A_c - B_c)$ represents the trust level in round three. The parameter d_c represents the speed of trust recovery due to trustworthy action, and Y_c represents the change in passing behavior between rounds six and seven due to an end-game effect.

We consider the opportunity for different communication conditions, c = 1 to 6, to influence the trust recovery parameters. We investigate the influence of the six different communication conditions that result from our 2 x 3 design. These conditions are the two deception conditions crossed by the three apology conditions depicted in Figure 2.

In our model, we construct parameter estimates as a function of both main effects and interaction terms for the communication conditions. We depict these in Table 1. Specifically, we consider a main effect m for each parameter, a main effect b_1 for deception, a main effect b_2 for promise, an effect b_3 for promise and apology, an interaction b_4 for deception and promise, and an interaction b_5 for deception, promise and apology. We create dummy variables to represent each condition: D_c = 1 for deception and 0 for no deception, E_c = 1 for promise and 0 for no promise, and F_c = 1 for apology and 0 for no apology. Using this framework we construct each parameter estimate as:

$$A_c = m_A + (D_c)b_{A1} + (E_c)b_{A2} + (E_c*F_c)b_{A3} + (D_c*E_c)b_{A4} + (D_c*E_c*F_c)\,b_{A5}$$

$$B_c = m_B + (D_c)b_{B1} + (E_c)b_{B2} + (E_c*F_c)b_{B3} + (D_c*E_c)b_{B4} + (D_c*E_c*F_c)b_{B5}$$

$$d_c = m_d + (D_c)b_{d1} + (E_c)b_{d2} + (E_c*F_c)b_{d3} + (D_c*E_c)b_{d4} + (D_c*E_c*F_c)b_{d5}$$

$$X_c = m_X + (D_c)b_{X1} + (E_c)b_{X2} + (E_c*F_c)b_{X3} + (D_c*E_c)b_{X4} + (D_c*E_c*F_c)b_{X5}$$

$$Y_c = m_Y + (D_c)b_{Y1} + (E_c)b_{Y2} + (E_c*F_c)b_{Y3} + (D_c*E_c)b_{Y4} + (D_c*E_c*F_c)b_{Y5}$$

For example, the long-term trust parameter, A_c, for the deception and promise alone condition (c=5) is a function of a main effect for deception (D_c=1), a main effect for promise alone (E_c=1), and an interaction between deception and promise (D_c*E_c=1):

$$A_5 = A_{Deception,Promise} = \quad m_A + b_{A1} + b_{A2} + b_{A4}$$

We obtain inferences from the model for parameter estimates, standard errors, and probability values using the Bayesian software package BUGS (Bayesian Inference Using Gibbs Sampling, http://www.mrc-bsu.cam.ac.uk/bugs), with uninformative priors for all parameters, while treating α_i as a random effect from a common Gaussian distribution. We use the Bayesian framework for two primary reasons. First, the distributions of interest may be skewed and we want an accurate assessment of standard errors, as compared to asymptotic ones obtained via classical maximum likelihood procedures. Second, since we want to make inferential statements regarding the "strength" of our hypothesized assertions, we use the Bayesian paradigm which allows for straightforward probability statements (Bayesian p-values) by counting the fraction of posterior draws supporting our hypotheses (Gelman, Meng, & Stern 1996).

We report results from posterior means obtained from running three independent chains of 15,000 draws each with the initial 10,000 draws of each chain discarded for burn-in. We assess convergence using the multiple F-test procedure of Gelman and Rubin (1992). Computing time for all three chains was roughly 0.15 seconds per iteration on a Dell 2.4 GHZ processing machine. The BUGS code used to implement our estimation is available from the authors upon request.

Post-decision survey. Immediately after making each passing decision, participants were asked how much they trust their partner (1: Not at all, 7: Completely). We examine these responses as a second dependent variable. These measures enable us to link perceptions and underlying motivations with actual behavior. To model these rating scores, we utilize a Gaussian distribution with a mean given by the *identical* functional form as the logit model in Equation (1). In this manner, we can directly compare inferences for both types of dependent variables.

Post-experiment survey. After participants received feedback from the final round of the experiment, they were asked to complete a two-page survey. This survey asked several questions related to their ex-post perceptions of trust. These questions asked participants about their perceptions of their partner in terms of their trust, integrity, honesty, and reliability (1: Not at all, 7: Completely). These measures were closely related, Cronbach's $\alpha = .8975$, and we use an average of these responses as our measure of ex-post trust. Participants were also asked demographic questions and open-ended questions regarding their perceptions of their counterpart's behavior in the experiment.

Results

A total of 184 participants completed the study. Just over half of the participants were male (53.3%), and almost all of our participants were between the ages of 19 and 22 (only 3 of 184 participants were over the age of 22). We considered gender differences in our models, and find no significant effects. As a result, we combine data across demographic variables for subsequent analysis.

Agreement Between Passing Decisions and Trust Ratings

We find very close agreement between passing decisions and trust ratings in our experiment. This was true across several types of analysis. First, we consider a random effects logistic regression for passing behavior, P_{irc}, modeled as a function of an individual parameter, α_i, an aggregate slope, β, and trust rating scores T_{irc} for each individual, *i*, each round, *r*, and each condition *c*.

$$P_{irc} = \text{logit}^{-1} (\alpha_i + \beta * T_{irc})$$

This model is highly predictive with trust rating parameter $\beta = 1.82$ (0.14); note that the coefficient for β is positive and large (thirteen standard errors away from 0). We conducted a second set of analyses to confirm that this relationship holds across individuals, with an individual slope parameter, β_i. Results from this model yield very similar results. In this case, the average β_i was 2.11 (0.26). In addition, the β_i parameter was significant for every participant; the least significant β_i parameter was 2.12 standard deviations above 0.

We also conducted a threshold analysis that provides a non-parametric view of the data. For each participant we examined the consistency between the trust ratings they provided when they passed and the trust ratings they provided when they took. Specifically, for each participant, we compared the maximum trust rating participants provided when they "Take" to the minimum trust rating they provided when they "Pass." We depict this formally. For each participant *i*, for rounds $r = 1$ to 6 and trust ratings T_{ir}, we calculate the following agreement score:

$$S_i = [\text{Max}_r\{T_{ir}|\text{Take}\} - \text{Min}_r\{T_{ir}|\text{Pass}\}] \quad (2)$$

We flag participants as lacking agreement with a fixed rating threshold over time if $S_i < 0$. This measure flags 16 participants. That is, only 16 of 184 participants provided a trust rating that was higher for *any* of the times they "Take" than the *minimum* they provided when they "Pass."

Even among these 16 participants, however, we find that disagreements are rare (typically happening only once), and that disagreements are small (typically by a single point).

We also conducted separate analysis fitting equation 1 for T_{irc} as the dependent variable. The model parameters for this model reflect the same pattern of results as those we find for the model representing passing decisions.

Taken together, these results suggest that passing decisions reflect underlying perceptions of trust. In our subsequent analysis we report results that use passing decisions as a behavioral representation of trust.

Modeling Passing Behavior

The focus of our analysis is on passing behavior, and in Figure 4 we depict actual passing behavior as the percentage of respondents passing by round across conditions. We fit our model (Equation 1) to the data, and find that our model of passing decisions closely tracks actual passing behavior. We report parameter estimates (posterior means) for each condition in Table 2 and depict the fitted model of trust recovery across conditions in Figure 5. The maximum deviation between the fitted and actual probabilities for any round is 6.80%, for the "Deception, No promise" condition in round 4, still a very close fit. For this model we set broad Gaussian priors for all parameters with s.d.=100 on the logit scale (extremely diffuse and uninformative), and we constrained $\delta > 0$.

Passing Behavior

We use the posterior draws from our model to compute the effects of each communication condition on passing decisions and the corresponding probabilities (Table 3) in each round. That is, Table 3 represents the posterior mean values of $\hat{P}_{irc}$ computed from the posterior draws obtained using BUGS. We use these estimated probability values as well as the parameter estimates from the model itself (Table 2) to assess the statistical significance of differences in passing rates across conditions. We define the cell entries in Table 3, which are differences in probabilities for various conditions by round as $\triangle P_{c,c'}(r)$, for differences between conditions c and c' in round r. For instance $\triangle P_{2,1}(3)$ represents the difference in trust between condition 2 (No Deception, Promise) and condition 1 (No Deception, No Promise) in round 3, which is the first round when the effect of the promise can be observed. Similarly, $\triangle P_{2,1}(8)$ represents the difference in long-term trust between condition 2 (No Deception, Promise) and condition 1 (No Deception, No Promise), which equals the long-term effect of a promise.

We first examine the influence of deception on the trust recovery process. We depict the effects of deception in the first row in Table 3 and in Figure 6. Supporting hypothesis 1, we find that for participants who received no other communication, deception significantly harms long-term levels of trust, $\triangle P_{4,1}(8) = -0.37$ (0.2), $p<.05$. That is, deception with no other communication leads to a 0.37 decrease on the probability scale of long-term passing. We also find that deception in round 2 increased passing, suggesting that the deceptive messages were initially effective in increasing passing behavior. We also note that after round 3, deception harms trust for each and every round including our hypothetical long-term round, i.e. as $r \geq 8$.

We next consider the influence of a promise on the trust recovery process. We depict the effects of a promise (with no other communication) in the second row in Table 3 and in Figure 7. We find that a promise significantly influenced early trust recovery, $\triangle P_{2,1}(3) = 0.579$ (0.1), $p<.001$, but that a promise did not significantly influence long-term trust recovery, $\triangle P_{4,1}(8) = 0.008$ (0.1), p=n.s. These findings support hypothesis 2a, but not hypothesis 2b. That is, we find that although a promise significantly speeded trust recovery, trustworthy actions alone are as effective in eventually restoring long-term trust as these same actions accompanied by a promise.

Surprisingly, we did not find a main effect for an apology in conjunction with a promise on the trust recovery process. We depict the effects of an apology coupled with a promise relative to a promise alone in the third row in Table 3 and in Figure 8. We do not find support for either hypothesis 3a or hypothesis 3b. An apology did not help to repair initial trust, $\triangle P_{3,2}(3) = -0.04$ (0.1), p=n.s., or long-term trust, $\triangle P_{3,2}(8) = -0.04$ (0.1), p=n.s. In the discussion section, we consider possible explanations for why the apology in this experiment did not significantly influence trust recovery.

We next consider the interaction between prior deception and a promise in restoring trust. We depict this interaction in the fourth row of Table 3 and in Figure 9 by comparing the difference between the deception and no deception conditions that either had or did not have a subsequent promise. We find a significant negative interaction in initial trust recovery, $\{\triangle P_{5,4}(3) - \triangle P_{2,1}(3)\} = -0.5$ (0.1), $p<.001$, but no significant interaction in long-term trust recovery, $\{\triangle P_{5,4}(8) - \triangle P_{2,1}(8)\} = 0.059$ (0.2), p=n.s. That is, prior deception harmed the initial effectiveness of a promise in restoring trust, but prior deception had no effect on the long-term influence of a promise on trust recovery. These findings support hypothesis 4a, but do not support hypothesis 4b.

We also examine the interaction between prior deception and an apology. We depict this interaction in the fifth row in Table 3 and in Figure 10 by comparing the difference between the

deception and no deception conditions that either had a promise and an apology or had a promise alone. We find that an apology increases relative long-term trust slightly more following deception than following no deception. These differences are marginally significant in rounds 5 and 6, but not significant long-term, $\{\triangle P_{6,5}(8) - \triangle P_{3,2}(8)\} = -.179$ (0.1), p=n.s. We find no effect on initial trust recovery, $\{\triangle P_{6,5}(3) - \triangle P_{3,2}(3)\} = -0.01$ (0.1), p=n.s. Taken together, we do not find support for either hypothesis 5a or hypothesis 5b.

Unrelated to our hypotheses, we find other expected patterns in our data. For example, repeated trustworthy actions significantly increased long-term trust; $\mu_B = 6.92$ (1.9), $p<.001$. Also, as expected, we find that participants passed significantly less often in the final, end-game round than they did in the penultimate round, $\mu_Y = -3.11$ (0.8), $p<.001$.

Economic Value of Communication

We next consider the economic and social welfare implications of communication. We use the passing probabilities and the even player decisions of "Return $0" for initial rounds and "Return $9" otherwise to compute the average earnings per round. We use actual passing probabilities for the initial rounds (rounds 1 and 2) and the trust recovery rounds (rounds 3 through 6) to compute average per round earnings. We also estimate average long-term earnings using parameter estimates for A_c in the passing model. For these values we estimate the long-term passing probability for each person i in condition c, P_{ic}, as:

$$P_{ic} = \text{logit}^{-1}(a_i + A_c) \quad (2)$$

and average across individuals. We report average earnings per round for both odd and even players in Table 4.

We first consider the economic implications of using deception for the deceiver. We find that while trustworthy actions restore trust and increase long-term earnings, trustworthy actions do not fully mitigate the harm caused by deception. While even players achieved short-term profits in the initial rounds with deception (with no other communication), earning $12.78 versus $8.10, even players earned less on average per round during the trust recovery process (rounds 3 through 6), $3.27 versus $5.10, and long-term, $4.65 versus $7.96.

We next consider the social welfare implications of both deception and trust restoring communication. The projected long-term earnings for both even and odd players combined following deception (and no other communication) are substantially lower than they are following no deception (and no other communication), $12.20 versus $16.61 per round. A subsequent promise and promise and apology, however, increase social welfare. Long-term, a promise and an

apology appear to increase social welfare only following deception. Following no deception, the long-term, per round combined earnings for each condition is close to the total potential earnings of $18. Following deception, however, the long-term per round combined earnings following no promise, a promise alone, and a promise and an apology were $12.20, $13.01, and $14.72. Thus, while an apology did not appear to have a significant impact on the parameter estimates, overall we see a moderate increase in social welfare from a promise and an apology versus a promise alone.

Final Survey

At the conclusion of the experiment, participants were asked a four-item trust inventory about their partner. Responses across these items were closely related, and we report the average rating across these items in Table 5. In a regression model, with final trust as the dependent variable and deception, a promise, and apology as independent variables, we find that final trust was significantly harmed by deception, significantly helped by a promise, but not significantly influenced by an apology. The standardized parameter estimates for deception, promise, and apology were –0.599 ($p<.001$), 0.253 ($p<.001$), and 0.010 (p=n.s.), respectively. These results offer a static, post-experiment perspective of trust that is consistent with our round-by-round analysis.

Discussion

While prior work has conjectured that trust is fragile and very difficult to repair, results from our investigation identify conditions under which trust is more or less likely to recover. Specifically, we find that trust can be effectively restored following a period of untrustworthy behavior as long as the untrustworthy behavior was not accompanied by deception. We find that deception causes significant and enduring harm to trust. In fact, deception harmed trust far more than untrustworthy actions alone.

We also identify a complicated relationship between promises and trust recovery. We had expected a promise to facilitate both initial and long-term trust recovery. Instead, we found that a promise helped initial trust recovery, but that long-term, trustworthy actions were as effective as trustworthy actions accompanied by a promise. We conjecture that a promise serves as a signal of intentions to change behavior. After a series of observed behaviors, however, the actions themselves effectively convey this same message.

Surprisingly, the apology in conjunction with a promise in this study did not restore trust more effectively than a promise alone. There are several potential reasons for why the apology in

this experiment was ineffective. First, our apology was always accompanied by a promise. In our case the promise alone may have signaled a change in intentions as effectively as a promise coupled with an apology. Alternatively, the type of apology we used in this experiment may not have been effective. Prior work has identified important attributes of apologies that our particular apology may have lacked. Specifically, our apology may not have been sufficiently substantial or sincere (Shapiro 1993). For example, our apology was not accompanied by an offer of penance (Bottom, Gibson, Daniels, & Murnighan, 2000). In addition, our apology may not have made a sufficiently clear attribution for prior behavior (Tomlinson, Dineen, & Lewicki 2002). Finally, our apology was written. Quite possibly, a written apology may have influenced participants less than an oral apology would have.

We also identify an important interaction between prior deception and a promise on trust recovery. Specifically, we find that prior deception harmed the initial effectiveness of a promise in restoring trust. In this case, deception may harm the trustee's credibility, and as a result subsequent promises may be viewed skeptically and discounted in the short-term.

By design, this experiment enables us to examine trust as a dynamic construct. Participants make decisions in a repeated game, and we focus our analysis of trust on passing decisions. We use passing decisions as our primary measure of trust for several reasons. First, passing decisions represent actual behavior and participants in our study had financial incentives to make these decisions carefully. Second, we believe that passing decisions in this experiment reflect trust decisions. We find very close agreement between passing decisions and our attitudinal measure of trust. In addition, when we fit a similar model for our attitudinal measure we find nearly identical results. Further, in the short essays participants wrote at the conclusion of the study, we found that participants were actively and strategically thinking about trust when they made their decisions.

Overall, trust recovery represents an important practical problem, and results from this work offer insight into the role actions, deception, and promises can play in changing trust over time. A number of important questions regarding the trust recovery process, however, remain. In particular, we made a number of choices in designing our experiment that afforded experimental control. A rich set of future studies could extend our understanding of trust recovery. For example, in our experiment we exposed participants to a consistent set of predetermined even player actions. In this case, participants learned about their counterpart's behavior consistently across conditions even if they did not pass. This enables us to provide common information across conditions and to isolate the effects of communication, but this aspect of our design favors trust

recovery. In some settings an untrustworthy episode may lead to relationship rupture, and subsequent trustworthy behavior will be more difficult to observe. In other settings, however, such as working with an untrustworthy boss or operating in an oligopoly setting (e.g. OPEC), people will observe subsequent actions even after an untrustworthy episode. As a result, while the common exposure to trustworthy actions affords experimental consistency and reflects some natural environments, the nature of trust recovery in other settings is likely to be more limited than we observe here.

Our design is also limited by our focus on anonymous relationships. This aspect of our design enables us to control for relationship effects across conditions, but future work should examine the trust repair process in richer contexts with mature relationships. According to Lewicki and Weithoff's (2000) conceptualization of trust relationships, trust violations in established relationships will lead to more severe consequences than those we observed in our early-stage relationships.

Our experiment was also constrained by the nature of our communication conditions. While this afforded consistency across participants, future work should examine a richer set of communication options. For example, future work should consider two-way communication, non-verbal communication, apologies alone, and contrast subtle, but potentially important differences between no communication when messages are allowed with no communication when messages are not allowed.

Future work could also extend our understanding of the interplay between communication and observed actions. For example, in our experiment, prior to round three our participants are influenced by round three messages as well as information about their counterparts' untrustworthy actions in round two. While we measure differences in passing rates in round three across conditions, future work could disentangle the effects of communication and observed actions within conditions.

In addition, future work should examine the relationship between the nature of the trust violation and the trust restoration process. For example, future work should explore the robustness of restored trust. Quite possibly, a second non-contiguous violation may harm trust far more than an initial violation. In a related vein, future work could examine the link between trust recovery and the nature of the trust betrayal. In our study, we can observe the relationship between a participant's trust betrayal experience, the number of times they trusted (passed) in early rounds and hence experienced untrustworthy behavior, and their trust recovery process. We did not,

however, manipulate participant's trust betrayal experience, and as a result we cannot draw causal inferences for this relationship. From our analysis it appears as if an individual's propensity to trust influences both initial- and late-stage behavior; participants who were trusting in early rounds (and experienced trust betrayal) were also more trusting in later rounds (and recovered trust more quickly). Future work should manipulate trust betrayal experiences to measure the effects of trust betrayal on the trust recovery process.

Another drawback of our experimental design is our use of deception. We gained experimental control and consistency within conditions, but there are important concerns about using deception in experiments. In fact, a substantial literature in social psychology has wrestled with the costs and benefits of using deception in experiments (c.f. Arndt, 1998; Hertwig & Ortmann, 2001). In many cases, the benefits of experimental control lead experimenters to use deception to investigate trust (e.g. Deutsch, 1958; Pillutla, Malhotra, & Murnighan, 2002; Malhotra & Murnighan, 2002) as well as many other topics (e.g. De Dreu, Carnevale, Emans & van de Vliert, 1994; Lim & Carnevale, 1995). In general, however, the decision to use deception should be made carefully and cautiously, and account for important externalities such as the future loss of experimenter credibility.

Overall, our results suggest that under some conditions trust can be regained quickly following a series of untrustworthy actions (e.g., no deception followed by a promise). This finding contradicts common assumptions regarding the trust recovery process and may inform practical prescriptions. For example, individuals should be careful not to make promises they cannot keep. Our results demonstrate that while trust can recover from a period of untrustworthy *actions*, deception causes significant and enduring harm. While deception may be tempting because it can be used to increase short-term profits for the deceiver, we find that the long-term costs of deception are very high. Our results also highlight the importance of a promise in speeding trust recovery. Importantly, a promise was not nearly as effective following deception as it was following no deception. We also found that trustworthy actions significantly, and in some cases dramatically, restore trust. Managers working to rebuild trust should be sure that people observe their trustworthy actions. Taken together, we find that when it comes to trust, actions matter, but they do not always speak louder than words.

References

Andreoni, J. & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. The Quarterly Journal of Economics, 116, 293-312.

Arndt, B. 1998. Deception can be acceptable, American Psychologist, 53, 805-806.

Ashraf, N., Bohnet, I., & Piankov, N. (2003). Decomposing trust. Working paper. Kennedy School of Government, Harvard University.

Atwater, L. (1988). The relative importance of situational and individual variables in predicting leader behavior. Group and Organization Studies, 13, 290-310.

Bazerman, M. (1994). Judgment in managerial decision making. New York: Wiley.

Berg, J., Dickhaut, J. & McCabe, K. (1995). Trust, reciprocity, and social history. Games and Economic Behavior, 10, 122-142.

Bottom, W., Daniels, S., Gibson, K. S., and Murnighan, J. K. (2002). When talk is not cheap: Substantive penance and expressions of intent in the reestablishment of cooperation. Organization Science, in press.

Bromiley, P., & Cummings, L. L. (1995). Transaction costs in organizations with trust. In R. J. Bies & B. Sheppard & R. J. Lewicki (Eds.), Research on negotiations in organizations, 5, 219-247. Greenwich, CT: JAI.

Buchan, N., Johnson, E., & Croson, R. (2002). Trust and reciprocity: An international experiment. Working paper, Wharton School, University of Pennsylvania.

Business Week. (1992). Sears gets handed a huge repair bill. September 14, 38.

Butler, J. (1999). Trust expectations, information sharing, climate of trust, and negotiation effectiveness and efficiency. Group and Organization Management, 24, 217-238.

Carr, A. (1968). Is business bluffing ethical? Harvard Business Review, 46, 143-153.

Charness, G. & Rabin, M. (2002). Understanding social preferences with simple tests. The Quarterly Journal of Economics, 117, 817-869.

Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment and personal need non-fulfillment. Journal of Occupational Psychology, 53, 39-52.

Cox, J. (2003). Trust and reciprocity: Implications of game triads and social contexts. Working paper, University of Arizona.

Dasgupta, P. (1988). Trust as a commodity. In D.G. Gambetta (Ed.), Trust: Making and Breaking Cooperative Relations. New York: Basil Blackwell.

Dawes, R. & Orbell, J. 1995. The benefit of optional play in anonymous one-shot prisoner's dilemma games. In Barriers to Conflict Resolution. Arrow, K., Mnookin, R., Ross, L., Tversky, A., & Wilson R., New York: Norton. 62-85.

DeDreu, C.K., Carnevale P. J. D., & Emans, B., van de Vliert, E. (1994). Effects of gain-loss frames in Negotiation: Loss aversion, mismatching, and frame adoption. Organizational Behavior and Human Decision Processes, 60, 90-107.

Deutsch, M. (1958). Trust and suspicion. Journal of Conflict Resolution, 2, 265-279.

Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. Human Relations, 12, 123-139.

Donaldson, T. (2001). The ethical wealth of nations. Journal of Business Ethics, 31, 25-36.

Farrell, J. & Rabin, M. (1996). Cheap talk. Journal of Economic Perspectives, 10, 103-118.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica, 6, 733-807.

Gelman, A. & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457-511.

Gibson, K., Bottom, W. & Murnighan, K. (1999). Once bitten: Defection and reconciliation in a cooperative enterprise. Business Ethics Quarterly, 9, 69-85.

Glaeser, E., Laibson, D., Scheinkman, J., & Soutter, C. (2000). Measuring Trust. The Quarterly Journal of Economics, 115, 811-846.

Gneezy, U. 2002. Deception: The role of consequences. Working paper, University of Chicago.

Hertwig, R. & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? Behavioral and Brain Sciences. 24, 383-451.

Hirsch, F. (1978). Social limits to growth. Cambridge, MA: Harvard University Press.

Ho, T. & Weigelt, K. (2002). Trust building among strangers. Working Paper, Wharton School, University of Pennsylvania 99-008.

Hosmer, L. T. (1995). Trust: The linking between organizational theory and philosophical ethics. Academy of Management Review, 20, 379-403.

Knack, S. & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. The Quarterly Journal of Economics, 112, 1251-1288.

Larrick, R. & Blount, S. (2002). Social context in tacit bargaining games. Working paper, Duke University.

Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.), Trust in organizations: Frontiers of theory and research. Thousand Oaks, CA: Sage.

Lewicki, R. J. & Wiethoff, C. (2000). Trust, trust development, and trust repair. In M. Deutsch & P. T. Coleman (Eds.), Handbook of Conflict Resolution: Theory and Practice. San Francisco, CA: Jossey-Bass.

Lim, R.G. & Carnevale, P.J. (1995). Influencing mediator behavior through bargainer framing. *International Journal of Conflict Management*, 6, 349-368.

Los Angeles Times. (1998). Hospital chain to pay $4.7 million. August 20, A16.

Malhotra, D. and Murnighan, J. K. (2002). The effects of formal and informal contracts on interpersonal trust. Administrative Science Quarterly, in press.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20, 709-734.

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.

O'Connor, K. & Carnevale, P. (1997). A nasty but effective negotiation strategy: Misrepresentation of a common-value issue. Personality and Social Psychology Bulletin, 23, 504-515.

Pillutla, M., Malhotra, D. and Murnighan, J. K. (2002). Attributions of trust and the calculus of reciprocity. Journal of Experimental Social Psychology, in press.

Ring, P. S., & Van de Ven, A. (1992). Structuring cooperative relationships between organizations. Strategic Management Journal, 13, 483-498.

Rousseau, M., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. Academy of Management Review, 23, 393-404.

Ross, W. & LaCroix, J. (1996). Multiple meanings of trust in negotiation research: A literature review and integrative model. International Journal of Conflict Management, 7, 314-360.

Rotter, J. (1971). Generalized expectancies for interpersonal trust. American Psychologist, 26, 443-452.

Sally, D. 1995. Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992. Rationality and Society, 7, 58-92.

Santoro, M. & Paine, L. (1993). Sears Auto Centers, Harvard Business School Case 9-394-010, Harvard Business School Publishing: Boston, MA.

Schurr, P. & Ozanne, J. (1985). Influence on exchange processes: Buyer's preconceptions of a seller's trustworthiness and bargaining toughness. Journal of Consumer Research, 11, 939-953.

Schweitzer, M. & Croson, R. (1999). Curtailing deception: The impact of direct questions on lies and omissions. The International Journal of Conflict Management, 10, 225-248.

Shapiro, D. (1991). The Effects of Explanations on Negative Reactions to Deceit. Administrative Science Quarterly, 4, 614-630.

Slovic, P. (1993). Perceived Risk, Trust, and Democracy. Risk Analysis, 13, 675-682.

Tomlinson, E., Dineen, B., & Lewicki, R. (2002). The road to reconciliation: The antecedents of reconciled trust following a broken promise. Working paper, Fisher College of Business, Ohio State University.

Thompson, L. (2001). The Mind and Heart of the Negotiator. Prentice Hall: Upper Saddle River, NJ.

Williams, M. (2001). In whom we trust: Group membership as an affective context for trust development. Academy of Management Review, 26, 377-396.

Table 1: Parameter Estimates

	No Trust Restoring Communication	*Promise*	*Promise and Apology*
No Deception	m	$m + b2$	$m + b2 + b3$
Deception	$m + b1$	$m + b1 + b2 + b4$	$m + b1 + b2 + b3 + b4 + b5$

Table 2: Parameter Estimates for Passing Behavior (Log Scale)

	Trust Recovery				Final Round
	Initial (X)	*Long-term (A)*	*Amount (B)*	*Speed (delta)*	*Decline (Y)*
Deception (D)	-3.63 (1.2) **	-3.89 (2.2) *	-3.95 (2.2) *	2.01 (2.8)	1.2 (1.1)
Promise (P)	2.77 (1.3) *	-0.51 (2.1)	-5.12 (2.1) ***	4.26 (2.7) †	0.29 (1.1)
Apology (A)	-1.24 (1.1)	-0.41 (1.3)	-0.08 (1.4)	-2.14 (4)	0.84 (1)
D, P	-0.78 (1.7)	0.95 (2.6)	4.88 (2.7) *	-3.41 (4.8)	-0.96 (1.5)
D, A	0.04 (1.5)	1.47 (1.8)	1.41 (1.9)	0.45 (5.2)	-1.13 (1.5)
Overall Mean	1.1 (1)	4 (1.9) *	6.92 (1.9) ***	0.77 (0.4)	-3.11 (0.8) ***

Log likelihood = -420.8
† p<.10, * p<.05, ** p<.01, *** p<.001

Table 3: Change in Passing Behavior by Round (Probability Scale)

	Round 2	*Round 3*	*Round 4*	*Round 5*	*Round 6*	*Long-term*
Deception, $\triangle P_{4,1}(r)$	0.357 (0.1) ***	-0.01 (0.1)	-0.16 (0.1) *	-0.3 (0.1) ***	-0.35 (0.1) ***	-0.37 (0.2) *
Promise, $\triangle P_{2,1}(r)$	0.128 (0.1) †	0.579 (0.1) ***	0.346 (0.1) ***	0.151 (0.1) **	0.072 (0.1)	0.008 (0.1)
Apology, $\triangle P_{3,2}(r)$	0.099 (0.1)	-0.04 (0.1)	-0.08 (0.04) †	-0.07 (0.04) †	-0.06 (0.1)	-0.04 (0.1)
D, P, $\triangle P_{5,4}(r) - \triangle P_{2,1}(r)$	-0.29 (0.1) *	-0.5 (0.1) ***	-0.21 (0.1) *	-0.04 (0.1)	0.025 (0.1)	0.059 (0.2)
D, A, $\triangle P_{6,5}(r) - \triangle P_{3,2}(r)$	0.021 (0.1)	-0.01 (0.1)	0.084 (0.1)	0.138 (0.1) †	0.162 (0.1) †	0.179 (0.1)

† p<.10, * p<.05, ** p<.01, *** p<.001

Table 4: Average Earnings per Round

	Odd Players				***Even Players***			
	Rounds 1-2	*Rounds 3-6*	*Long-Term* †		*Rounds 1-2*	*Rounds 3-6*	*Long-Term* †	
No D, No P	$3.30	$7.70	$8.65		$8.10	$5.10	$7.96	
No D, P	$3.00	$8.56	$8.68		$9.00	$7.69	$8.04	
No D, P&A	$2.51	$8.37	$8.57		$10.46	$7.11	$7.71	
D, No P	$1.74	$7.09	$7.55		$12.78	$3.27	$4.65	
D, P	$2.10	$7.40	$7.75		$11.70	$4.20	$5.26	
D, P&A	$1.84	$7.50	$8.18		$12.48	$4.50	$6.54	

† Long-Term earnings are estimated, assuming $P_{ic} = \text{logit}^{-1}(a_i + A_c)$

Table 5: Average Post-Experiment Trust†

	No Promise	*Promise*	*Promise & Apology*
Deception	2.21 (0.98)	2.65 (1.12)	2.64 (1.10)
No Deception	3.48 (1.36)	4.77 (1.37)	4.85 (1.09)

† These values represent the average of four questions:
Q1. How much do you trust your partner? (1: Completely, 7: Not at all)
Q2. How much integrity do you think your partner has? (1: A great deal, 7: None at all)
Q3. How honest do you think your partner was? (1: Completely, 7: Not at all)
Q4. How reliable do you think your partner is? (1: Very reliable, 7: Not at all reliable)

Figure 1: Trust Game

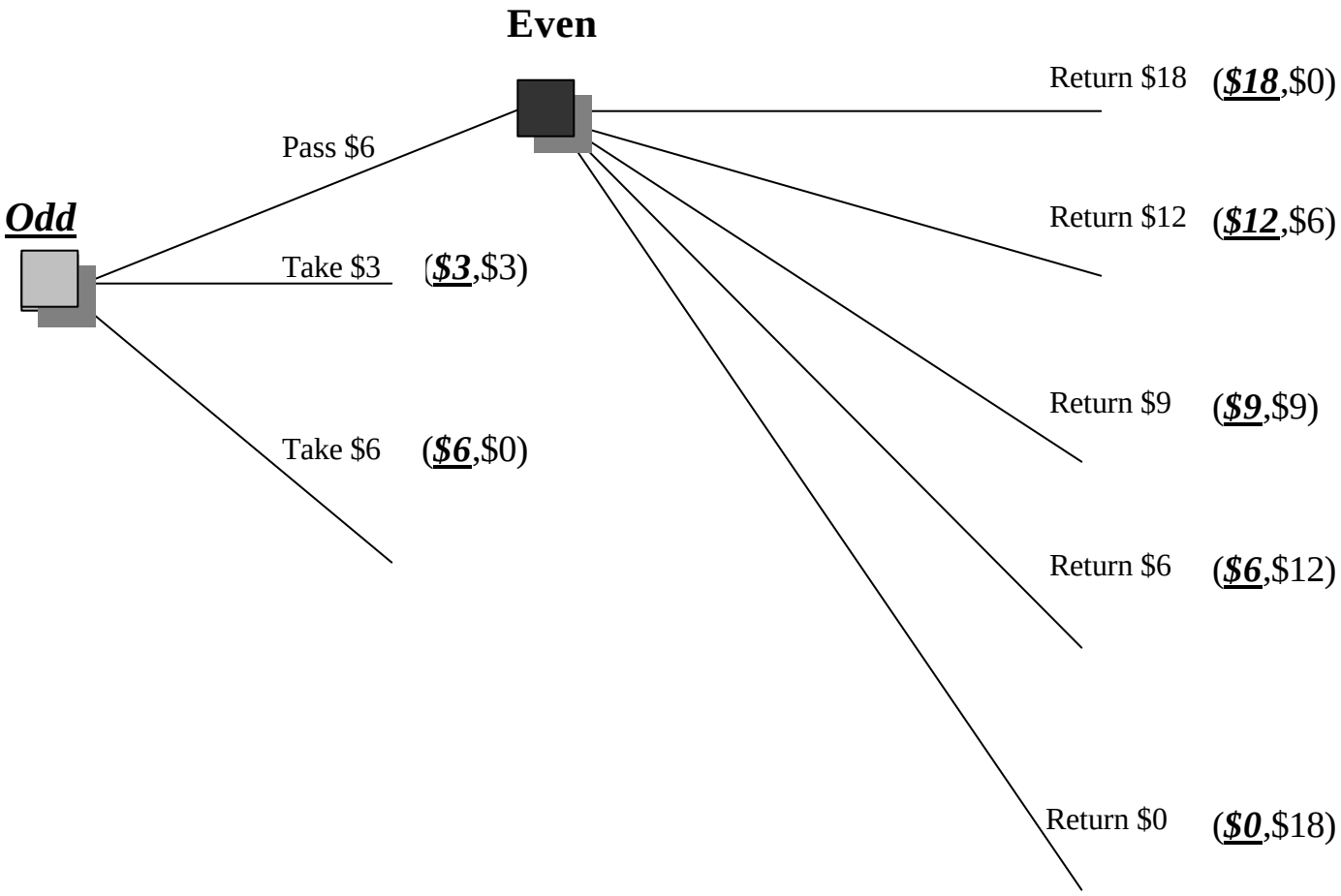


Figure 2: Experimental Design

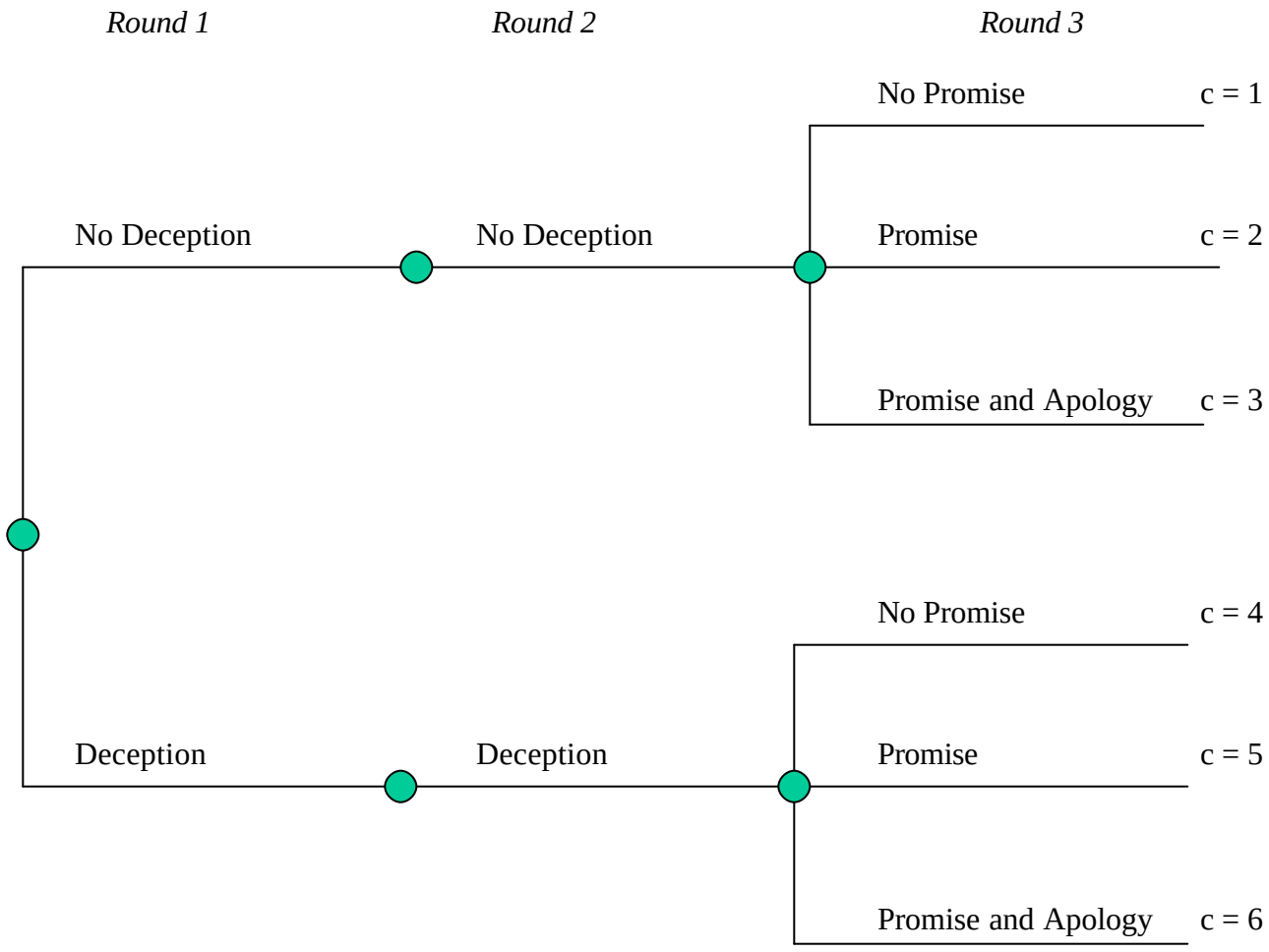


Figure 3: Trust Recovery Model

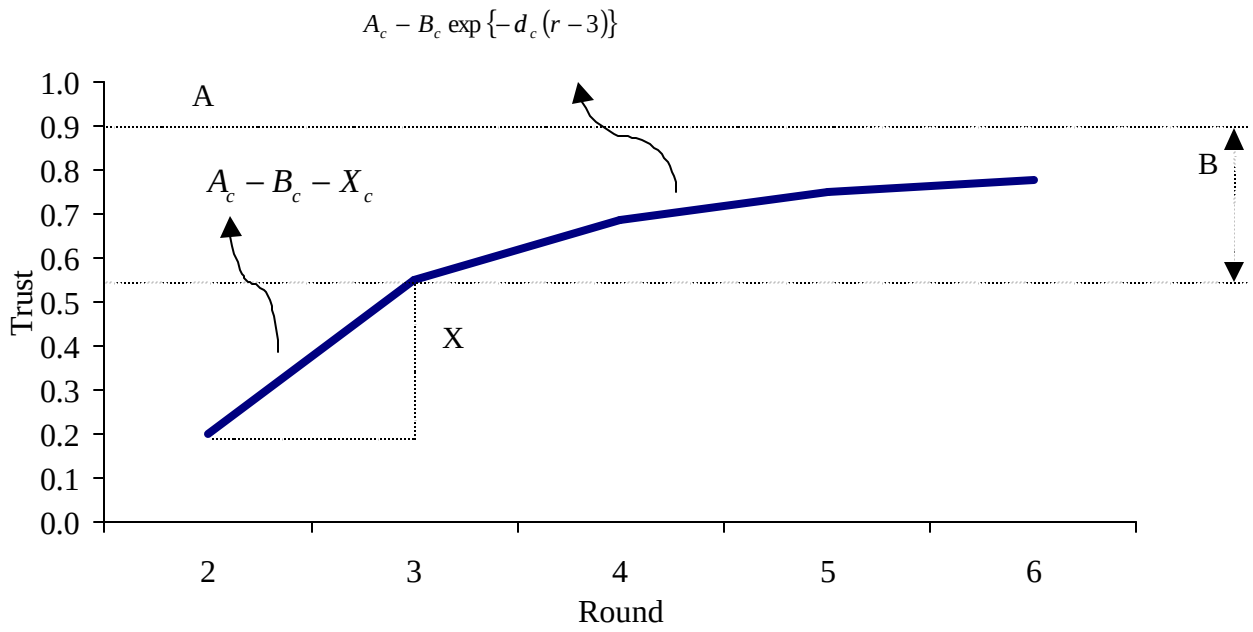


Figure 4: Passing Decisions by Conditions (Actual)

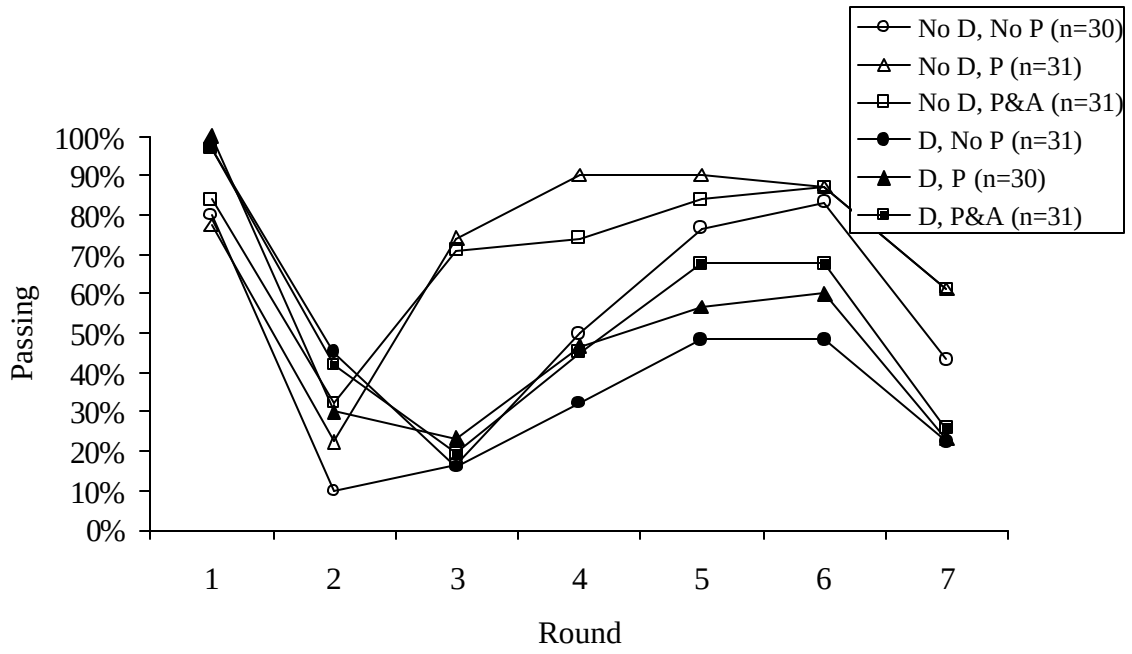


Figure 5: Passing Decisions by Conditions (Model Fit)

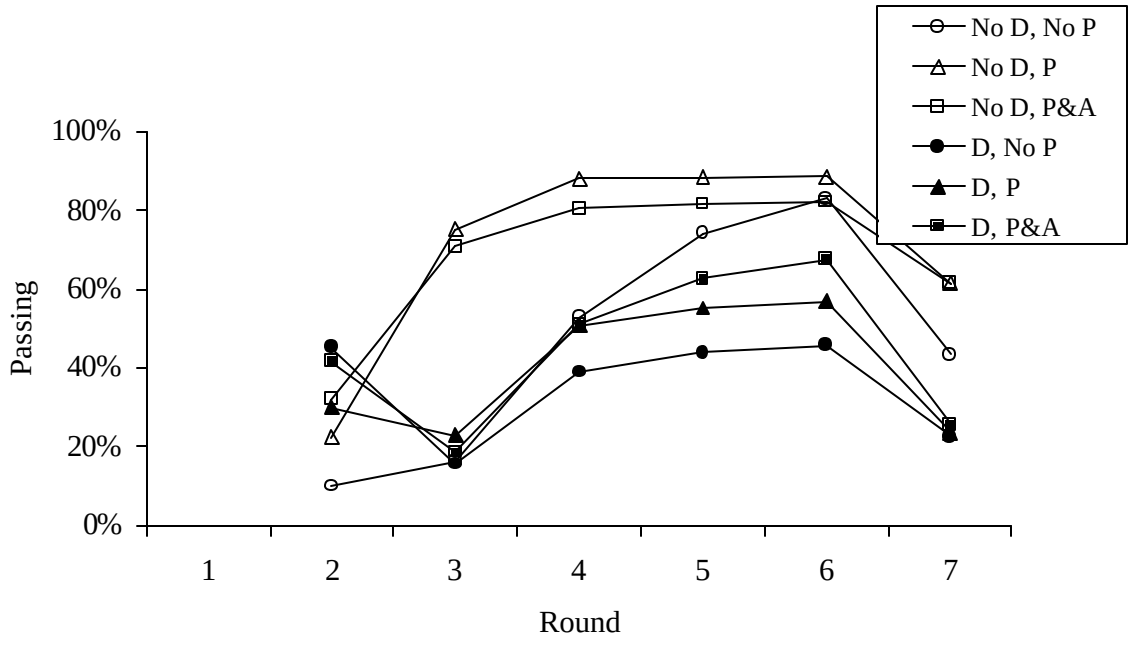
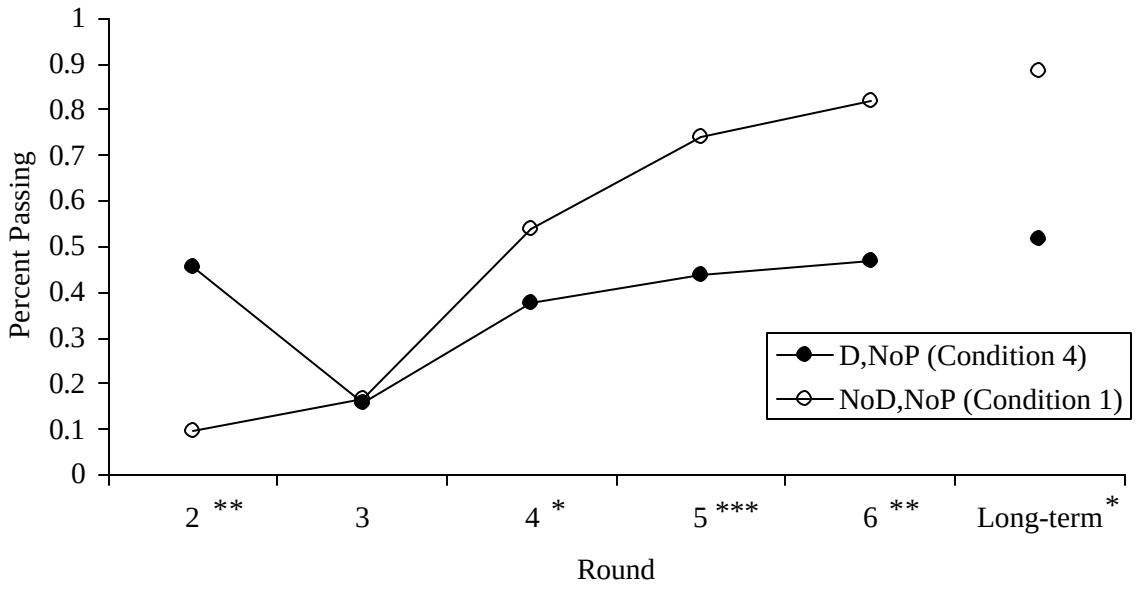


Figure 6: Deception and Trust Recovery: Fitted Values for $\triangle P_{4,1}(r)$



For Figures 6 through 10 the significance of differences in each round are indicate by † p<.10, * p<.05, ** p<.01, *** p<.001

Figure 7: A Promise and Trust Recovery: Fitted Values for $\triangle P_{2,1}(r)$

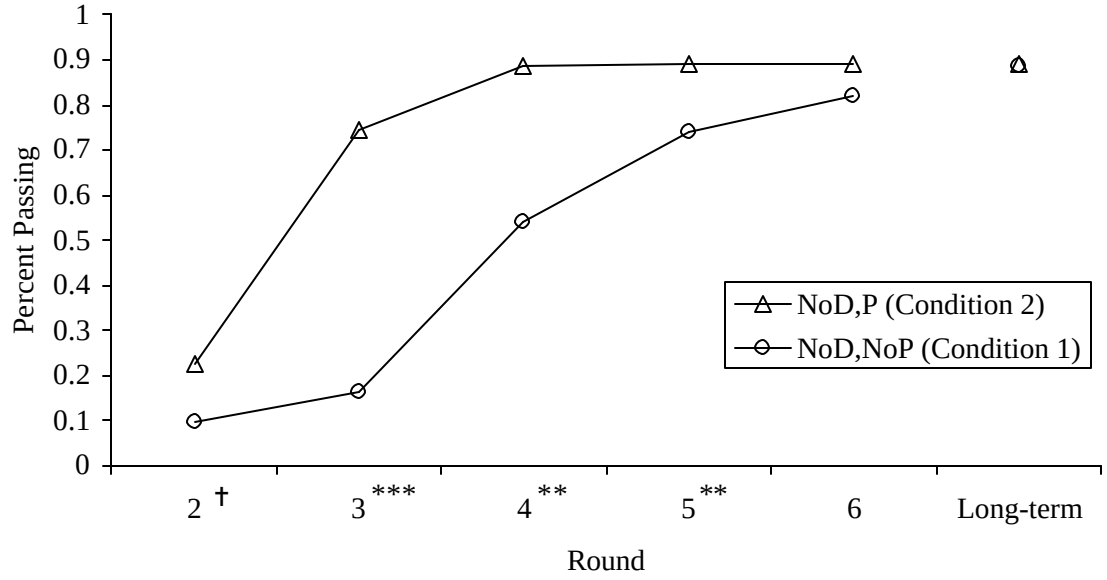


Figure 8: An Apology and Trust Recovery: Fitted Values for $\triangle P_{3,2}(r)$

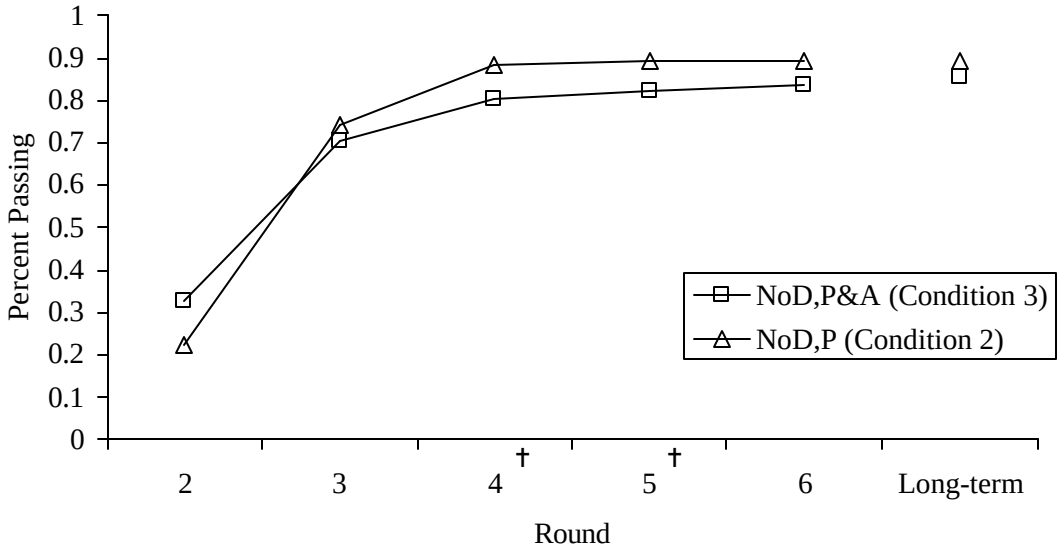


Figure 9: Deception and a Promise on Trust Recovery: Fitted Values for $\triangle P_{5,4}(r)$ - $\triangle P_{2,1}(r)$

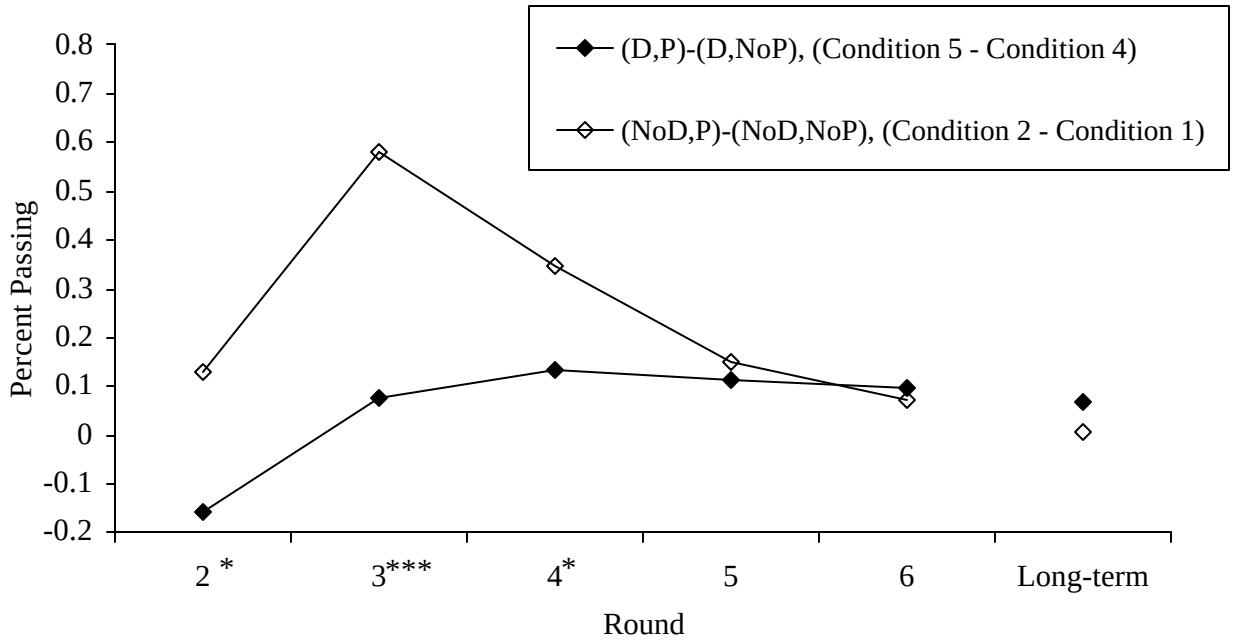


Figure 10: Deception, a Promise, and an Apology on Trust Recovery: Fitted Values for $\triangle P_{6,5}(r)$ - $\triangle P_{3,2}(r)$