PROMISES AND LIES: RESTORING VIOLATED TRUST

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

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

Eric T. Bradlow
Wharton School, University of Pennsylvania
761 JMHH, Marketing and Statistics
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 Roy Lewicki and Keith Murnighan for helpful comments, and Samantha Rudolph and Clifford Lou for research assistance.

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 behavior can significantly speed the trust recovery process, but prior deception harms the effectiveness of a promise in accelerating trust recovery.

PROMISES AND LIES: RESTORING VIOLATED TRUST

Trust is essential for organizations (Donaldson, 2001; Lewicki, McAllister, & Bies, 1998). Trust enables managers to lead more effectively (Atwater, 1988, Bazerman, 1994) and negotiate more efficiently (Valley, Moag & Bazerman, 1998). At the same time, however, we know that trust is often violated. Trust violations can range from serious misdeeds that constitute fraud (Santoro & Paine, 1993, Business Week, 1992, Los Angeles Times, 1998) to more common forms of trust violations, such as the use of deception in negotiations (Boles, Croson, & Murnighan, 2000; Carr, 1968; O'Connor & Carnevale, 1997; Schweitzer & Croson, 1999; Steinel & De Dreu, 2004).

Surprisingly little, however, is known about the consequences of violating trust. While common wisdom presumes that trust violations can cause severe relationship damage (e.g., Slovic, 1993), little work has 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 observe how trust is harmed over time in the presence of deception and untrustworthy behavior, and how a promise, an apology, or a promise and an apology repair trust when combined with trustworthy actions.

BACKGROUND AND HYPOTHESES

Trust

A number of definitions of trust have been advanced, and in this work 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.

A substantial literature has identified a number of individual and contextual factors that influence trust (Dasgupta, 1988; Deutsch, 1960; Lewicki & Weithoff, 2000; Mayer, Davis, & Schoorman, 1995; Ross & LaCroix, 1996; Williams, 2001). Much of this work has identified perceptions of concern as a key driver of trust judgments. For example, managers are trusted more when they demonstrate interest in their team members' ideas (Korsgaard et al., 1995). Related work has identified favorable attributions for past behavior as essential for trust development (Larrick & Blount, 2002; Pillutla, Malhotra, & Murnighan, 2002). In fact, Malhotra and Murnighan (2002) found that the use of binding contracts can actually harm trust development, because subjects who used binding contracts make situational rather than personal attributions for trustworthy behavior.

Surprisingly, most prior research has examined trust as a static construct (e.g., Glaeser, Laibson, Scheinkman, & Soutter, 2000). Only a few studies have considered how trust develops (see Pillutla, Malhotra, & Murnighan, 2003), and little extant research has considered how trust might recover after it has been harmed. 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. Similarly, Slovic (1993) postulates that lost trust can take a long time to rebuild and that in some cases, lost trust may never be restored. In an experimental study, Kim, Ferrin, Cooper, and Dirks (2004) examined differences in trust following apologies and denials for allegations of improper behavior. They asked participants to assume the

role of a manager in charge of a hiring decision and to judge the trustworthiness of a potential candidate. Participants watched video-taped interviews of a hypothetical candidate who was accused of filing an incorrect tax return in her previous job. The candidate either denied or apologized for the infraction. Kim et al. (2004) found that apologies led to higher trust judgments than denials when the candidate was accused of incompetence, but that denials led to higher trust judgments when the candidate was accused of a breach of integrity.

Related research has examined cooperation in prisoners dilemma and social dilemma games. In a study involving a repeated prisoners dilemma game, Gibson, Bottom, and Murnighan (1999) examined methods to restore cooperation following uncooperative behavior. They found that apologies and offers of penance were effective in reestablishing cooperation. In another study, Buchan, Johnson, and Croson (2002) found that non-task communication increased trust. Other work investigating the dynamics of cooperation has found that cooperation levels are higher when participants are members of the same social network (Bowles & Gintis, 2004), such as from the same neighborhood; when participants have the ability to punish a free riding counterpart (Fehr & Gachter, 2000); and when participants initiate an interaction by cooperating with each other (Clark & Sefton, 2001).

Present Investigation

In this paper, we use experimental methods to investigate changes in trust behavior over time. Our work differs from prior investigations in several important ways. No prior work has examined trust recovery from the perspective of the trustor, and no prior work has examined the long-term effects of deception on trust. In this work, we

disentangle the harmful effects of untrustworthy behavior from deception, and we describe the interaction between deception and subsequent promises and apologies in rebuilding trust.

Consistent with Bok's (1982) work, we define deception as intentional acts of deceit. In our experiment, we expose participants to written statements that are disconfirmed by subsequent actions. We define untrustworthy behavior as actions that would harm a vulnerable trustee, and we operationalize untrustworthy behavior in a setting with economic incentives. In this work, we consider trust restoration from the perspective of the target of prior untrustworthy behavior.

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 that we used 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 player 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 dominated by the decision to take $6, but affords participants who do not trust their counterpart an opportunity to make an altruistic choice and give their partner some money.

All of our participants make decisions as odd players, and receive feedback and prepared messages from even player confederates. In our experiment, every participant is exposed to a consistent set of even player actions regardless of their passing or taking decisions. That is, 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 set of actions. Prior work has documented the tendency of people to initiate interactions with high levels of trust (Ekman, 1996; McKnight, Cummings, & Chervany 1998; Meyerson, Weick & Kramer, 1996), and we expect initial untrustworthy actions (the even player returns $0) to harm trust, and subsequent trustworthy actions (the even player returns $9) to restore trust.

Hypotheses

In our experiment, some participants receive messages from their counterparts. Our hypotheses, summarized in Table 1, focus on the effects of this communication on the trust recovery process. By examining trust in a repeated game, we are able to measure changes in trust over time.

A strong null hypothesis predicts no effect for any communication. All of the communication in our experiment constitutes "mere" or "cheap" talk. Participants know that it is costless for their counterpart 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.

In our study, participants both observe behavior and receive messages. Rather than discounting messages as merely cheap talk, we postulate that participants will rely

upon observed behavior, messages, and the interplay between observed behavior and messages to gauge the trustworthiness of their counterpart. This prediction is consistent with prior work that has found that cheap talk can influence behavior (Buchan, Johnson, & Croson, 2002; Croson, Boles & Murnighan, 2003).

We expect messages to influence participants' assessments of their counterpart in several ways. Prior work has demonstrated that the attributions individuals make for behavior influence trust judgments (Larrick & Blount, 2002; Pillutla, Malhotra, & Murnighan, 2002). Messages communicate important information about intentionality, and we expect messages and the correspondence between messages and observed behavior to influence the attributions participants make about their counterpart's past behavior. We also expect the correspondence between messages and observed behavior to influence participants' assessments of the credibility of their counterpart. Prior work has demonstrated that credibility directly impacts judgments of trustworthiness (Kim et al., 2004).

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. Thus, 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.

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 (Bok, 1982). In this experiment, the combination of deception and untrustworthy behavior clarifies the untrustworthy acts as intentional. As a result, we expect participants exposed to deception and untrustworthy behavior to judge the likelihood that their counterpart is untrustworthy as higher than participants exposed to untrustworthy behavior alone. Although we expect subsequent trustworthy acts to increase perceptions of trustworthiness, we do not expect subsequent trustworthy acts to fully mitigate the harmful effects of deception. That is, 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, untrustworthy actions combined with deception will decrease the long-term level of trust more than the same untrustworthy actions without deception.

Trust Restoring Communication and Trust Recovery

Next, we examine the effects of trust-restoring communication on trust recovery. Prior work suggests that damaged trust may be very difficult to repair. In general, the trust recovery process is assumed to be slow and incomplete (Slovic, 1993, Lewicki &

Bunker, 1996). We consider the role of a promise, an apology, and both a promise and an apology, in conjunction with trustworthy actions, in rebuilding trust.

Consistent with prior work that has found that promises facilitate cooperation (Orbell, Dawes, & Kragt, 1988; Rubin & Brown, 1975; Schlenker, Helm & Tedeschi, 1973), we expect promises to facilitate the trust restoration process. Untrustworthy actions may be multiply determined, and trustors' attributions for untrustworthy behavior may be labile and subject to impression management. Morrison & Bies (1991: 523) define impression management as "a motive to control how one appears to others." Impression management can be both defensive, an attempt to avoid creating an unfavorable image, and assertive, an attempt to create a favorable impression (Tedeschi & Melburg, 1984; Tedeschi & Norman, 1985). Promises represent an assertive impression management approach designed to convey positive intentions about future acts. If believed, promises are likely to restore positive expectations about future behavior and to improve subjective assessments regarding the likelihood that an individual is a trustworthy type of person. 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 participants 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 repair initial trust more than no trust restoring communication.

Hypothesis 2b: In the absence of other communication, a promise to change behavior will increase the long-term level of trust more than no trust restoring communication.

We next consider the effect of an apology. We adopt Schlenker and Darby's (1981: 271) definition of an apology as an "admission of blameworthiness and regret for an undesirable event." Prior work has found that apologies influence judgments about transgressors. Specifically, prior work has found that respondents rate transgressors who apologize more favorably and as less culpable than they rate transgressors who do not apologize (Ohbuchi, Kameda, & Agarie, 1989; Schwartz et al., 1978; Ohbuchi & Sato, 2001; Darby & Schlenker, 1982). We expect apologies to influence favorably the assessments participants make about their counterpart's type.

Schlenker & Darby (1981) identify five key components of an apology: (1) statement of apology (e.g., I'm sorry), (2) expressions of remorse (e.g., I feel badly), (3) offer of restitution, (4) self-castigation (e.g., I was an idiot), and (5) a request for forgiveness. In addition to these components, we consider other key elements of an apology: (6) a promise regarding future behavior, and (7) an explanation for the transgression.

In this work, we disentangle the effects of a promise from other components of an apology, and we consider an apology that includes three primary apology components: a statement of apology, an expression of remorse, and self-castigation. Our apology read, "I really screwed up, I shouldn't have done that. I'm very sorry I tried taking so much these

last two rounds." Our apology did not include an offer of restitution, a request for forgiveness, a promise, or an explanation. We postulate that an apology will increase subjective judgments regarding the likelihood that an individual is a trustworthy type of person, and in this setting, we test whether the three components we included in our apology are sufficient to restore short-term or long-term trust.

Hypothesis 3a: In the absence of other communication, an apology will repair initial trust more than no trust restoring communication.

Hypothesis 3b: In the absence of other communication, an apology will increase the long-term level of trust more than no trust restoring communication.

Interaction between Promise and Apology

We expect both promises and apologies to increase subjective perceptions of trustworthiness by conveying information about a counterpart's underlying nature or "type." We expect an apology coupled with a promise to restore trust more quickly and more completely than either a promise or an apology alone. However, we conceptualize promises and apologies as partial substitutes. That is, assuming main effects for a promise alone and an apology alone, we expect the cumulative effect of a promise combined with an apology to be less than the sum of the independent effects. This will show up as a negative interaction.

Hypothesis 4a: We predict a negative interaction between a promise and an apology in repairing initial trust.

Hypothesis 4b: We predict a negative interaction between a promise and an apology in increasing long-term levels of trust.

Interaction between Deception and a Subsequent Promise

We consider the interaction between a promise and prior deception. In the short term, we expect promises to restore trust more following no deception than following deception. When a promise follows deception, the trustor is unlikely to perceive the promise as credible, and the promise is likely to be significantly discounted. Promises can articulate positive expectations regarding future behavior, but messages that are not credible will fail to change impressions and expectations (Tedeschi & Reiss, 1981). As a result, a promise that follows deception is far less likely to increase the subjective likelihood that the trustee is a trustworthy type of person than is a promise that does not follow deception.

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. We expect trustworthy actions to restore overall trust and to build credibility in the promise in both the deception and no deception conditions. We do not expect trustworthy actions to restore trust any more for participants who were deceived and who received a promise than for participants who only received a promise. Both in the short run and in the long run, we expect a promise to be more effective in restoring trust when participants were not deceived than when participants were deceived.

Hypothesis 5a: Prior deception will harm the initial effectiveness of a promise in restoring

trust.

Hypothesis 5b: Prior deception will harm the long-term effectiveness of a promise in

restoring trust.

Interaction between Deception and a Subsequent Apology

We consider the interaction between an apology and prior deception. For an apology to be effective, receivers need to perceive the apology as sincere (Shapiro, 1991). When an apology follows deception, however, it is likely to be significantly discounted. In the short term, we expect an apology to restore trust more fully following no deception than following deception.

The long-term effects of an apology following deception, relative to an apology following no deception, will depend on the extent to which trustworthy actions restore credibility in the apology and restore overall trust. As with a promise, we expect trustworthy actions to exert a strong effect on general perceptions of trustworthiness and to build credibility in the apology in both conditions. Overall, we expect an apology to be more effective in restoring trust when participants were exposed to untrustworthy actions and not deceived than when participants were exposed to untrustworthy actions and deceived.

Hypothesis 6a: Prior deception will harm the initial effectiveness of an apology in restoring

trust.

Hypothesis 6b: Prior deception will harm the long-term effectiveness of an apology in

restoring trust.

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, which asks participants to specify the strategies they would use before they learn about their counterpart's actual decision. In the Appendix, we provide an excerpt from our instructions that explains this aspect of our design to our participants.

The strategy method allows odd players to learn even players' contingent decisions, regardless of odd players' actual decisions in that round. In our case, the decisions communicated to participants were identical within each round.

In our experiment, after each round, every odd player learns what his or her even player counterpart chose. In our study, every even player counterpart chooses to return \$0 the first two rounds and to return \$9 for rounds three through seven. That is, every participant in our experiment 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 trustworthy actions each participant observes 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 *un*trustworthy actions. In the second phase, involving the middle four rounds (r = 3 through 6), all participants

observed trustworthy actions. We added a third phase, the final round ($r = 7$), to account for a potential end-game effect.

Sample and Materials

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.

Upon arrival to the experiment, participants were randomly separated into two different rooms. Within each room, participants were randomly assigned to a treatment condition and a pairing number. (We collected data for this study in two separate time periods. In the second time period, we collected data for the following two conditions: No Deception, No Promise, Apology and Deception, No Promise, Apology.)

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 eight between-subject communication conditions that we depict in Figure 2. These conditions result from a 2x2x2 design: two deception conditions in rounds 1 and 2 (Deceptive messages prior to rounds 1 and 2, No messages prior to rounds 1 and 2) and two promise and two apology conditions in round 3 (No message prior to round 3, Promise alone prior to round 3, Apology 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.

Four promise and apology conditions dictated communication prior to round 3. In the *promise and apology condition* the round 3 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 *apology alone 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." In the *no promise-no apology 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.

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.

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 collected survey responses. After each of the seven rounds, we asked participants, "How much do you trust your partner?" (1: Completely Trust, 7: Do Not Trust at All). 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, reciprocity, fairness, and altruism (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 with an ANOVA structure would require 56 parameters to model passing decisions for each of the eight 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 8; 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. a promise) 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 8, to influence the trust recovery parameters. We investigate the influence of the eight different communication conditions that result from our 2 x 2 x 2 design. These conditions are the two deception conditions crossed by the two promise and two 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 2. We obtain inferences from the model for parameter estimates, standard errors, and tail probability values (p-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. Bayesian models using BUGS have been developed to study a wide range of phenomena from longitudinal biomedical and health data (Guo & Carlin, 2004) to bidding behavior at online auction sites (Park and Bradlow, 2005).

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 $\alpha = .902$, 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 262 participants completed the study. Just over half of the participants were male (52.3%), and almost all of our participants were between the ages of 19 and 22 (only 16 of 262 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 $\beta = 1.99$ (SE = 0.14, t = 14.2, $p<.001$); note that the coefficient for β is positive and large (fourteen 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 2.43 (SE = 0.36, t =

6.8, $p<0.001$). In addition, the β_i parameter was significant for every participant; the *least* significant β_i parameter was 2.06 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 24 participants. That is, only 24 of 262 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 24 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 3 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 5.9%, for the "No Deception, Promise, Apology" condition in round 4, still a very close fit.

Passing Behavior

In Table 3, we represent the posterior mean values of $\hat{P}_{irc}$ computed from the posterior draws obtained using the BUGS software. We use the posterior draws from our model to compute the effects of each communication condition on passing decisions and the corresponding probabilities in each round. We define the cell entries in Table 4, 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, No Apology) and condition 1 (No Deception, No Promise, No Apology) 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, No Apology) and condition 1 (No Deception, No Promise, No Apology), which equals the long-term effect of a promise.

We use data from Table 4 to test our hypotheses. As depicted in Table 2, we consider the main effects of our three conditions, deception, promise, and apology, as well as the three two-way interactions of these effects.

First, we examine the influence of deception on the trust recovery process. We depict the effects of deception in Figure 6 and in the first row in Table 4. Supporting hypothesis 1, we find that for participants who received no other communication, deception significantly harms long-term levels of trust, $\triangle P_{5,1}(8) = -0.37$ (SE=0.2, t=1.9, $p<.05$). 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 $r ? 8$.

Second, we 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 4 and in Figure 7. We find that a promise significantly influenced early trust recovery, $\triangle P_{2,1}(3) = 0.579$ (SE=0.1, t=5.8, $p<.001$), but that a promise did not significantly influence long-term trust recovery, $\triangle P_{2,1}(8) = 0.008$ (SE=0.1, t=0.08, 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.

Third, we consider the effects of an apology on the trust recovery process. We depict the effects of an apology (with no other communication) in the third row in Table 4 and in Figure 8. we find that an apology did not significantly influence either early trust

recovery, $\triangle P_{3,1}(3) = 0.06$ (SE=0.07, t=.86, p=n.s.) or long-term trust recovery, $\triangle P_{3,1}(8) = 0.02$ (SE=0.1, t=0.2, p=n.s.). That is, we do not find support for hypotheses 3a or 3b. In this study, we find that trustworthy actions alone are as effective in eventually restoring long-term trust as these same actions accompanied by an apology. 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 three two-way interaction hypotheses represented in Table 1. We first examine the interaction between a promise and an apology, and depict the effects of this interaction in the fourth row in Table 4 and in Figure 9. Though in the expected direction, we find no significant interaction between a promise and an apology in repairing initial trust, $\triangle P_{4,3}(3) - \triangle P_{2,1}(3) = -0.11$ (SE=0.11, t = -1.0, p=n.s.), or long-term trust, $\triangle P_{4,3}(8) - \triangle P_{2,1}(8) = -0.06$ (SE=0.12, t = -0.5, p=n.s.). That is, we do not find support for either hypothesis 4a or hypothesis 4b. Of course, this is not surprising given the lack of an effect for an apology alone.

The second interaction we examine is the interaction between prior deception and a promise on restoring trust. We depict this interaction in the fifth row of Table 4 and in Figure 10 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, $\{\triangle P_{6,5}(3) - \triangle P_{2,1}(3)\} = -0.51$ (SE=0.11, t=-4.96, p<.001), but no significant interaction in long-term trust recovery, $\{\triangle P_{6,5}(8) - \triangle P_{2,1}(8)\} = 0.08$ (SE=0.2, t= .4, 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 5a, but do not support hypothesis 5b.

The third interaction we examine is the interaction between prior deception and an apology on restoring trust. We depict this interaction in the sixth row in Table 4 and in Figure 11 by comparing the difference between the deception and no deception conditions that either had or did not have a subsequent apology. We find no interactions between deception and an apology on either initial trust recovery $\{\triangle P_{7,5}(3) - \triangle P_{3,1}(3)\}$ = -0.07 (SE=0.10, t=-0.7, p=n.s.)} or long-term trust recovery $\{\triangle P_{7,5}(8) - \triangle P_{3,1}(8)\}$ = 0.04 (SE=0.22, t= .18, p=n.s.)}. This finding is consistent with our earlier finding that apology alone had no effect either initially or long-term.

We summarize the results of our hypotheses tests in Table 7. This table notes the significant harmful effects of deception on long-term levels of trust, the significant beneficial effects of a promise on initial levels of trust, and the negative interaction between prior deception and a promise on initial levels of trust. Overall, we find no significant effects for our apology.

Unrelated to our hypotheses, we find other expected patterns in our data. For example, repeated trustworthy actions significantly increased long-term trust. In our model, μ_B denotes the difference between trust levels in round 3 and trust levels long-term. We find that observing trustworthy actions significantly repairs trust; μ_B =6.92 (SE=1.9, t=3.65, 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. In our model, μ_Y denotes the difference in trust between rounds 6 and 7. We find a significant end game effect on trust; μ_Y = -3.11 (SE=0.8, t=-3.89, 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 5.

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 $12.10 versus $8.53, even players earned less on average per round during the trust recovery process (rounds 3 through 6), $3.95 versus $6.48, and long-term, $5.49 versus $8.01.

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), $12.28 versus

$16.72 per round. A subsequent promise or apology, however, slightly increases social welfare, especially following deception. Following no deception, the long-term, per round combined earnings for each condition is close to the total potential earnings of $18. These values range from $16.26 to $16.99. Following deception, however, the long-term per round combined earnings for odd and even players was $12.28 with no promise and no apology, $13.24 with a promise only, $12.99 with an apology only, and $14.80 with a promise and an apology. In our setting, each percentage increase in the passing rate translates into a $0.12 increase in social welfare. As a result, social welfare differences across communication conditions can reflect relatively small and insignificant differences in passing rates.

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 ratings across these items in Table 6. We ran a regression model, with final trust as the dependent variable and deception, a promise, an apology, and interaction terms as independent variables. The model was significant, $F(7,254)=26.83$ ($p<.001$; adj. R-square=.41), and we find that final trust was significantly harmed by deception $\beta = -1.27$ (SE = 0.3, $t = -4.22$, $p<.001$), significantly helped by a promise $\beta = 1.28$ (SE = 0.3, $t = 4.25$, $p<.001$), but not significantly influenced by an apology $\beta = 0.51$ (SE = 0.29, $t = 1.78$, $p=.08$). The only significant interaction was between deception and a promise $\beta = -0.84$ (SE = 0.43, $t = -1.97$, $p =.05$). 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 challenge and qualify this claim. 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 trust harmed by deception never fully recovers. Unlike untrustworthy actions alone, untrustworthy actions combined with deception causes enduring harm to trust.

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.

We found that an apology did not facilitate trust recovery. Prior work has documented effects for an apology, and we consider differences between our study and prior work as well as characteristics of our apology to reconcile this discrepancy.

First, respondents in most prior studies did not experience an actual transgression (Kim et al., 2004; Schwartz et al., 1978; Ohbuchi & Sato, 2001; Darby & Schlenker, 1982). Prior work asked respondents to consider hypothetical scenarios in which a transgressor either did or did not apologize. These observers were then asked to make judgments about the transgressor. In our study, participants actually experienced a transgression. Whereas social desirability may have influenced respondents when judging

the culpability of a transgressor in a scenario study, the actual experience may have led to different reactions.

In Ohbuchi, Kameda, and Agarie's (1989) study, participants did experience an actual transgression. In their study, participants perform poorly on an aptitude test because of the mistakes a confederate made in administering it. The harm in this study resulted from threats to self-esteem or embarrassment for having performed poorly. The confederate's apology includes the claim that she is "solely responsible for the subject's poor performance" (p.220). In this case, the apology itself mitigates the harm of the transgression, because it makes clear that the threat to self-esteem was entirely fabricated. This mitigation is a form of restitution, an apology component not present in our study.

Second, most prior studies asked respondents to provide general impressions about the transgressor (e.g., "Do you think Pat is a good person" Darby & Schlenker, 1982 p. 745). In our study, we measured participants' willingness to rely upon the transgressor in a specific context with monetary stakes. An apology may influence some impressions (e.g., liking), but leave trust judgments unchanged.

Third, our experimental context is different from other studies in that participants in our study faced a repeated game. In most prior work (e.g., Darby & Schlenker, 1982; Ohbuchi, Kameda, & Agarie, 1989; Ohbuchi & Sato, 2001), participants were asked to evaluate a single transgression and were not asked to consider relying upon the transgressor in the future. The transgressor's future actions in our study are likely to be particularly salient to our participants when reading an apology. In fact, we manipulated claims about future behavior explicitly with our promise condition. Notably, Kim et al. (2004) did ask participants to consider hiring a hypothetical employee who had

committed a transgression in a previous job. In their study, their apology included a statement regarding future intentions; the apology "admitted responsibility for the trust violation, apologized for the infraction, and *stated that such an incident would not happen again*" (emphasis added, p. 108). In our study, we disentangle an apology from a promise. We find a significant effect for a promise, but no effect for our form of apology alone.

Fourth, we crafted our apology to be general enough to apply across both deception and no deception conditions. Quite possibly, a specific apology (e.g., an apology that directly addresses the use of deception) may be effective in mollifying a specific transgression (e.g., deception).

Fifth, other characteristics of our apology may have limited its effectiveness. For example, our apology may not have been sufficiently long or sincere (Schlenker & Darby, 1981; Shapiro, 1993). More specifically, our apology lacked an explanation for the transgression (Goffman, 1971; Hodgins, Liebeskind, & Schwartz, 1996; Tomlinson, Dineen, & Lewicki, 2004), it lacked a request for forgiveness, and it lacked an offer of restitution (Bottom, Gibson, Daniels, & Murnighan, 2000). Had we used a different type of apology or delivered the same apology in a different way (e.g., orally rather than in writing), perhaps an apology would have been effective in restoring trust.

We also 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 be discounted.

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, at least retrospectively, claimed to be 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, promises, and apologies 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, such as employee- management interactions (Wiesenfeld, Brockner & Martin, 1999; Wiesenfeld, Brockner & Thibault, 2000). 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, and contrast subtle, but potentially important differences between no communication when messages are allowed with no communication when messages are not allowed. Prior work has found that the amount of communication (Kim, 1997) and even the communication medium itself (Valley, Moag & Bazerman, 1998; Valley et al., 2002) can significantly influence trust and the efficiency of bargaining outcomes, and future work should explore the interplay between specific messages and the communication medium.

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 participants' 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).

The nature of the existing relationship is also likely to influence the trust recovery experience following trust betrayal. Quite possibly, an apology may be more important in resorting trust in an established relationship than it is in restoring trust in an emerging relationship. Future work should examine different types of relationships and measure the effects of different types of betrayals on the trust recovery process.

Trust betrayal is likely to induce intense negative emotions. Recent work has begun to document the relationship between emotions and trust (Dunn & Schweitzer, 2005), and future research investigating trust recovery should explore the role of emotions. We postulate that trust repair tactics that mitigate negative emotions will be more effective in restoring trust than other, similarly informative strategies, that do not mitigate these emotions.

In our experimental design we used 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, the decision to use deception should be made carefully and cautiously.

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

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.

Bok, S. (1982). Secrets. New York: Pantheon.

Boles, T., Croson, R., & Murnighan, K. (2000) Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes, 83*: 235-259.

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.

Bowles, S. & Gintis, H. (2004). Persistent parochialism: trust and exclusion in ethnic networks, Journal of Economic Behavior and Organization, 55, 1-23.

Bromiley, P., & Cummings, L. L. (1995) Transaction costs in organizations with trust. In R. J. Bies and B. Sheppard and 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.

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

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

Clark, K. & Sefton, M. (2001). The Sequential Prisoner's Dilemma: Evidence on Reciprocation, Economic Journal, 111: 51-68.

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

Croson, R., Boles, T., & Murnighan, J. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior and Organization, 51*: 143-159.

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

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.

Dunn, J. & Schweitzer, M. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology,* 88: 736-748.

Ekman, P. (1996). Why Don't We Catch Liars? *Social Research*, 63: 801-817.

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

Fehr, E. & Gachter, S. (2000). Cooperation and Punishment in Public Goods Experiments. The American Economic Review, 90: 980-994.

Goffman, E. (1971) Relations in Public, Penguin, Harmondsworth.

Guo, X. & Carlin, B.P. (2004) Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58: 16--24.

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. *Quarterly Journal of Economics, 115*: 811-846.

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

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

Hodgins, H. S., Liebeskind, E., & Schwartz, W. (1996). Getting out of hot water: Facework in social predicaments. *Journal of Personality and Social Psychology*, 71: 300-314.

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

Kim, P. (1997). Strategic timing in group negotiations: The implications of forced entry and forced exit for negotiators with unequal power. *Organizational Behavior and Human Decision Processes. 71*: 263-286.

Kim, P., Ferrin, D., Cooper, C., & Dirks, K. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence- versus integrity-based trust violations. *Journal of Applied Psychology, 89*: 104-118.

Korsgaard, M., Schweiger, D., & Sapienza, H. (1995) Building commitment, attachment, and trust in strategic decision-making teams: The role of procedural justice. *Academy of Management Journal, 38*: 60-84.

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 and 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 and 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., & Murnighan, J. K. (2002) The effects of formal and informal contracts on interpersonal trust. *Administrative Science Quarterly, 47*: 534-559.

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

McKnight, D., Cumings, L., & Chervany, N. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, 20: 23, 473-490.

Meyerson, D., Weick, K., & Kramer, R. (1996). Swift trust and temporary groups. In R. Kramer & T. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166-195). Thousand Oaks, CA: Sage.

Morrison, E. & Bies, R. (1991) Impression management in the feedback-seeking process: A literature review and research agenda, *Academy of Management Review*, 16: 522-541.

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.

Orbell, J., Dawes, R., & Kragt, A. van de (1988). Explaining discussion induced cooperation. Journal of Personality and Social Psychology, 54: 811-819.

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.

Park, Y.H. & Bradlow, E.T. (2005) An integrated model for bidding behavior in internet auctions: Whether, who, when, and how much. *Journal of Marketing Research*, 42, 470-482.

Pillutla, M., Malhotra, D., & Murnighan, J. K. (2003) Attributions of trust and the calculus of reciprocity. Journal of Experimental Social Psychology, 39: 448-455.

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. and 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.

Rubin, J. & Brown, B. (1975). The social psychology of bargaining and negotiation. New York, NY: Academic Press.

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

Schlenker, B. R., & Darby, B. W. (1981). The use of apologies in social predicaments. *Social Psychology Quarterly*, 44: 271-278.

Schlenker, B. R., Helm, B., & Tedeschi, J. T. (1973). The effects of personality and situational variables on behavioral trust. *Journal of Personality and Social Psychology*, 25: 419-427.

Schweitzer, M., & Croson, R. (1999) Curtailing deception: The impact of direct questions on lies and omissions. *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.

Steinel, W. & De Dreu, C. (2004). Social Motives and Strategic Misrepresentation in Social Decision Making. *Journal of Personality and Social Psychology, 86*: 419-434.

Tedeschi, T., & Melburg, V. (1984). Impression management and influence in the organization. *Research in the Sociology of Organizations*, 3: 31-58.

Tedeschi, J. & Norman, N. (1985). Social power, self-presentation, and the self. In B. Schlenker (Ed.) *The Self and Social Life*, (p. 293–322). New York: McGraw-Hill.

Tedeschi, James & Reiss, M. (1981). Verbal strategies in impression management, In C. Antaki (Ed.), *The Psychology of Ordinary Explanations of Social Behaviour*. (p. 156-187). New York: Academic Press, 156–187.

Tomlinson, E., Dineen, B., & Lewicki, R. (2004) The road to reconciliation: Antecedents of victim willingness to reconcile following a broken promise. *Journal of Management*, *30*: 165-188.

Valley, K. L., Moag, J., & Bazerman, M. H. (1998) 'A matter of trust': Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior and Organization, 34*: 211-238.

Valley, K. L., Thompson, L., Gibbons, R., & Bazerman, M. H. (2002) How communication improves efficiency in bargaining games. *Games and Economic Behavior, 38*: 127-155.

Wiesenfeld, B. M., Brockner, J., & Martin, C. (1999) A self-affirmation analysis of survivors' reactions to unfair organizational downsizing. *Journal of Experimental Social Psychology, 35*: 441-460.

Wiesenfeld, B. M., Brockner, J., & Thibault, V. (2000) Procedural fairness, managers' self-esteem, and managerial behaviors following a layoff. *Organizational Behavior and Human Decision Processes, 83*: 1-32.

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

Appendix

The following is an excerpt from the instructions we gave the odd players:

How Each Round Works
There are $6 at the start of each round. The Odd player makes the first decision. The Odd player can “Take $6”, “Take $3”, or “Pass $6”.

As the Odd player, …
if you choose “Take $6”, you will earn $6 and the Even player will earn $0.
if you choose “Take $3”, you will earn $3 and the Even player will earn $3.
if you choose “Pass $6”, the amount of money grows to $18, and the Even player decides how much of the $18 to return to you.

The Even player can “Return $18”, “Return $12”, “Return $9”, “Return $6”, or “Return $0”.
If the Even player chooses “Return $18”, you earn $18 and Even earns $0.
If the Even player chooses “Return $12”, you earn $12 and Even earns $6.
If the Even player chooses “Return $9”, you earn $9 and Even earns $9.
If the Even player chooses “Return $6”, you earn $6 and Even earns $12.
If the Even player chooses “Return $0”, you earn $0 and Even earns $18.

In each round, both Odd and Even players are asked to indicate what they would do. Note that the game may actually end earlier so the Even player's choice may not influence the outcome of the game. For example, if an Odd player chooses to “Take $6,” then the choice of the Even player is meaningless (since the game ends—with Odd earning $6 and Even earning $0). Only when the Odd player chooses to “Pass $6” will the choice of the Even player matter. Still, in each round both Odd and Even players will record their decisions.

Table 1: Summary of Hypotheses

	Hypothesized Effect on Trust Recovery	
	Short Term	*Long Term*
Deception	--	H1: Negative
Promise	H2a: Positive	H2b: Positive
Apology	H3a: Positive	H3b: Positive
Interaction of Promise & Apology	H4a: Negative	H4b: Negative
Interaction of Deception & Promise	H5a: Negative	H5b: Negative
Interaction of Deception & Apology	H6a: Negative	H6b: Negative

Table 2: Parameter Estimates

	Condition	*Estimate*
No Deception		
No Promise, No Apology	1	μ
Promise, No Apology	2	$\mu + \beta_1$
No Promise, Apology	3	$\mu + \beta_2$
Promise, Apology	4	$\mu + \beta_1 + \beta_2 + \beta_3$
Deception		
No Promise, No Apology	5	$\mu + \beta_4$
Promise, No Apology	6	$\mu + \beta_1 + \beta_4 + \beta_5$
No Promise, Apology	7	$\mu + \beta_2 + \beta_4 + \beta_6$
Promise, Apology	8	$\mu + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7$

Table 3: Parameter Estimates for Passing Behavior (Log Scale)

	Trust Recovery				Final Round
	Initial (X)	*Long-term (A)*	*Amount (B)*	*Speed (delta)*	*Decline (Y)*
Deception (D)	-3.56 (1.36) **	-4.02 (2.95) †	-4.24 (2.89) *	2.06 (2.75)	1.34 (1.07)
Promise (P)	2.89 (1.39) *	-0.61 (2.4)	-5.46 (2.38) **	4.13 (2.77)	0.31 (1.08)
Apology (A)	9.58 (5.67) **	0.6 (3.13)	-0.08 (3.07)	-0.09 (0.66)	-1.62 (1.12) †
P&A	-10.91 (5.79) **	-0.99 (3.94)	0.17 (3.91)	-1.81 (4.11)	2.57 (1.51) *
D, P	-0.98 (1.79)	1.18 (3.5)	5.41 (3.45) *	-3.53 (4.79)	-1.09 (1.48)
D, A	-8.34 (5.77) *	-0.23 (4.04)	0.56 (3.95)	-0.67 (3.56)	1.27 (1.48)
D, P&A	8.49 (5.98) *	1.64 (5.32)	0.5 (5.24)	0.82 (6.41)	-2.43 (2.06)
Overall Mean	1.1 (1.09)	4.37 (2.21) *	7.4 (2.18) ***	0.745 (.51)	-3.25 (.8) ***

Log likelihood = -571.3

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

Table 4: Differences in Probability by Round (Probability Scale)

	Round 2		*Round 3*		*Round 4*		*Round 5*		*Round 6*		*Long-term*	
Deception, $\Delta P_{5,1}(r)$	0.35 (.08)	***	0.0 (.07)		-0.15 (.08)	†	-0.30 (.06)	***	-0.36 (.07)	***	-0.37 (.17)	*
Promise , $\Delta P_{2,1}(r)$	0.13 (.07)	†	0.58 (.08)	***	0.35 (.06)	***	0.15 (.05)	**	0.07 (.05)		0.00 (.08)	
Apology , $\Delta P_{3,1}(r)$	-0.10 (.04)	*	0.06 (.07)		0.07 (.07)		0.04 (.05)		0.02 (.06)		0.02 (.10)	
P, A, $\Delta P_{4,3}(r) - \Delta P_{2,1}(r)$	0.19 (.10)	*	-0.11 (.11)		-0.16 (.09)	†	-0.10 (.06)	†	-0.09 (.07)		-0.06 (.12)	
D, P, $\Delta P_{6,5}(r) - \Delta P_{2,1}(r)$	-0.28 (.12)	*	-0.51 (.11)	***	-0.23 (.10)	*	-0.04 (.07)		0.03 (.09)		0.08 (.20)	
D, A, $\Delta P_{7,5}(r) - \Delta P_{3,1}(r)$	-0.07 (.10)		-0.07 (.10)		-0.05 (.11)		0.01 (.07)		0.03 (.10)		0.04 (.22)	
D, P, A, $[\Delta P_{8,7}(r) - \Delta P_{6,5}(r)] - [\Delta P_{4,3}(r) - \Delta P_{2,1}(r)]$	0.09 (.17)		0.07 (.15)		0.14 (.15)		0.13 (.10)		0.14 (.14)		0.14 (.29)	

† p<.10, * p<.05, ** p<.01, *** p<.001
Note: Conditions 1 through 8 correspond to the framework depicted in Table 2.

Table 5: Average Earnings per Round

Condition		***Odd Players***			†	***Even Players***			†
	No Deception	*Rounds 1-2*	*Rounds 3-6*	*Long-Term*		*Rounds 1-2*	*Rounds 3-6*	*Long-Term*	
1	*No Promise, No Apology*	$3.10	$7.60	$8.68		$8.30	$5.20	$8.04	
2	*Promise, No Apology*	$2.76	$8.49	$8.69		$9.24	$7.77	$8.07	
3	*No Promise, Apology*	$3.85	$7.67	$8.75		$6.00	$5.71	$8.24	
4	*Promise, Apology*	$2.37	$8.23	$8.57		$10.60	$7.26	$7.70	
	Deception								
5	*No Promise, No Apology*	$1.60	$7.04	$7.57		$12.92	$3.31	$4.71	
6	*Promise, No Apology*	$2.10	$7.35	$7.81		$11.70	$4.25	$5.43	
7	*No Promise, Apology*	$2.19	$7.12	$7.75		$11.19	$3.58	$5.24	
8	*Promise, Apology*	$1.74	$7.35	$8.20		$12.58	$4.65	$6.60	

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

Table 6: Average Post-Experiment Trust†

Condition			
	No Deception	*Average Trust (s.d.)*	
1	*No Promise, No Apology*	3.48	(1.36)
2	*Promise, No Apology*	4.77	(1.43)
3	*No Promise, Apology*	3.99	(1.38)
4	*Promise, Apology*	4.85	(1.43)
	Deception		
5	*No Promise, No Apology*	2.21	(1.47)
6	*Promise, No Apology*	2.65	(1.42)
7	*No Promise, Apology*	2.31	(1.47)
8	*Promise, Apology*	2.64	(1.42)

† 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)

Table 7: Summary of Results

	Trust Recovery Results			
	Short Term		*Long Term*	
Deception	--		**H1: Negative**	*
Promise	**H2a: Positive**	***	H2b: Positive	
Apology	H3a: Positive		H3b: Positive	
Promise & Apology	H4a: Negative		H4b: Negative	
Deception & Promise	**H5a: Negative**	***	H5b: Negative	
Deception & Apology	H6a: Negative		H6b: Negative	

Figure 1: Trust Game

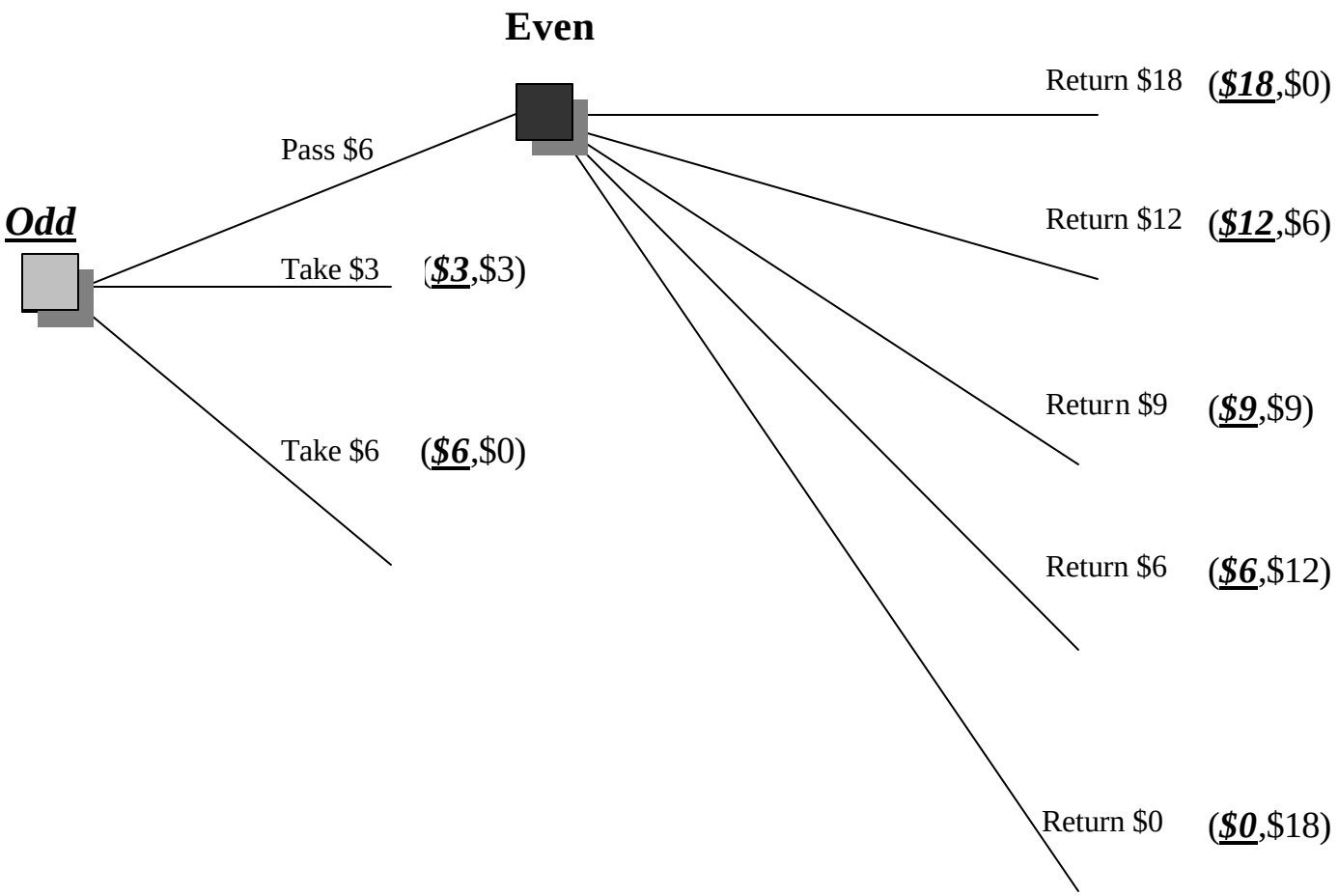


Figure 2: Experimental Design

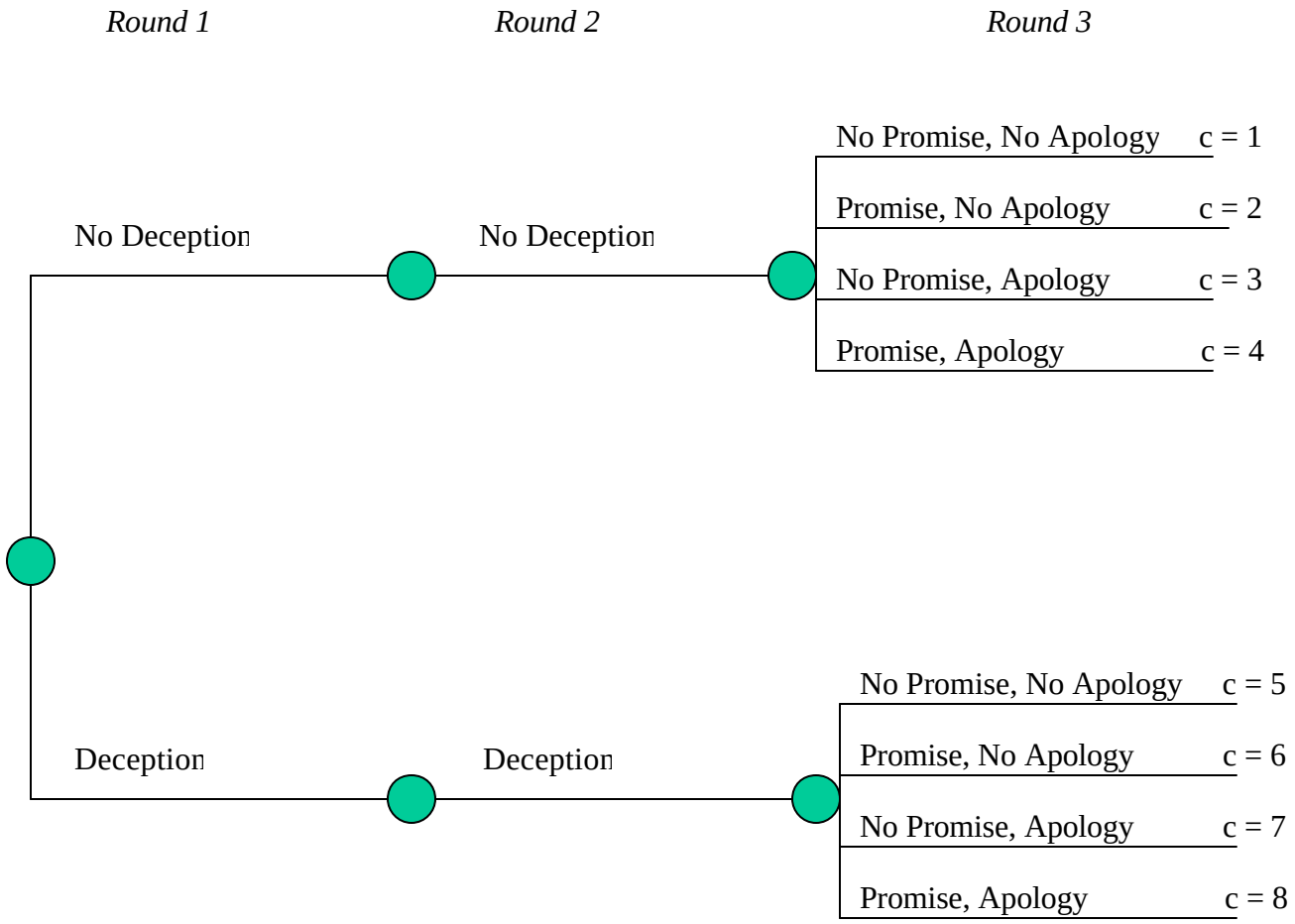


Figure 3: Trust Recovery Model

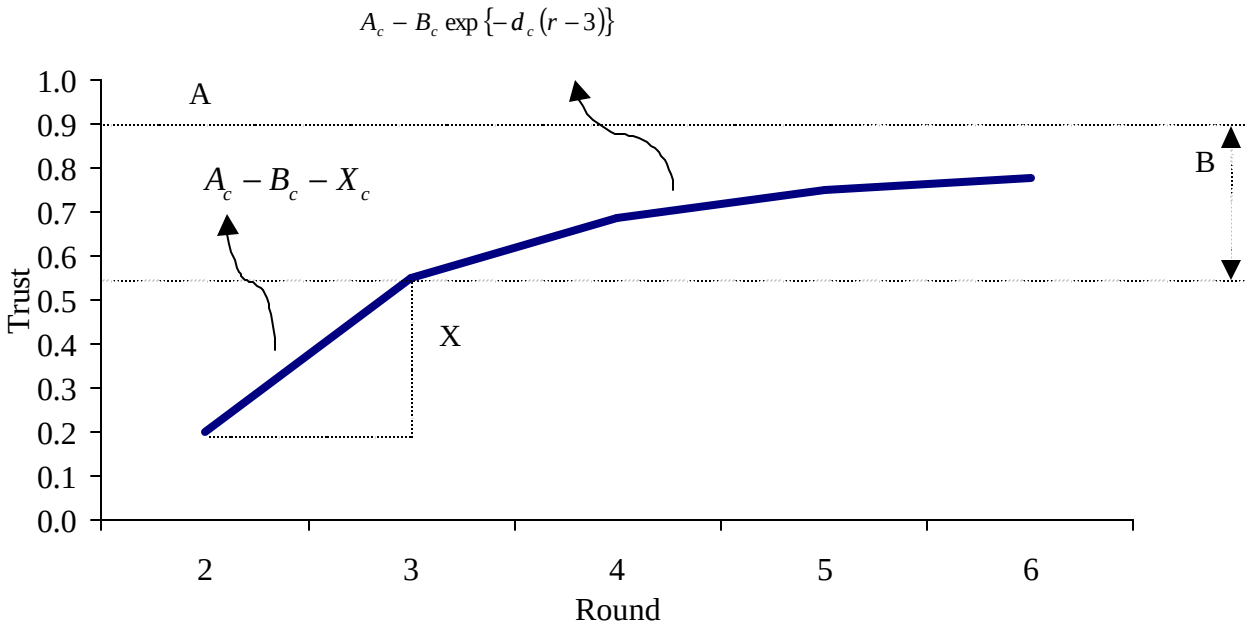


Figure 4: Passing Decisions by Conditions (Actual)

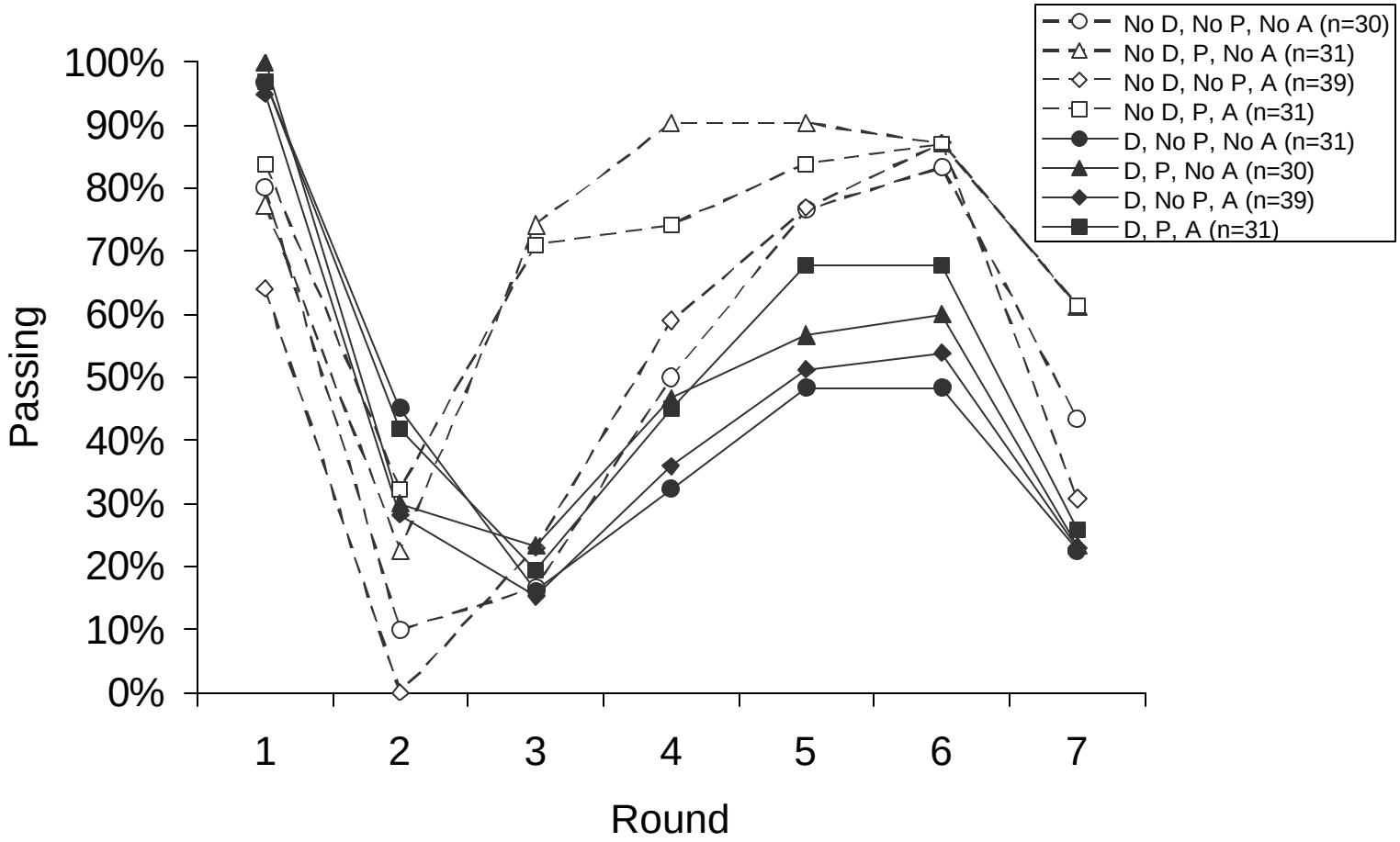


Figure 5: Passing Decisions by Conditions (Model Fit)

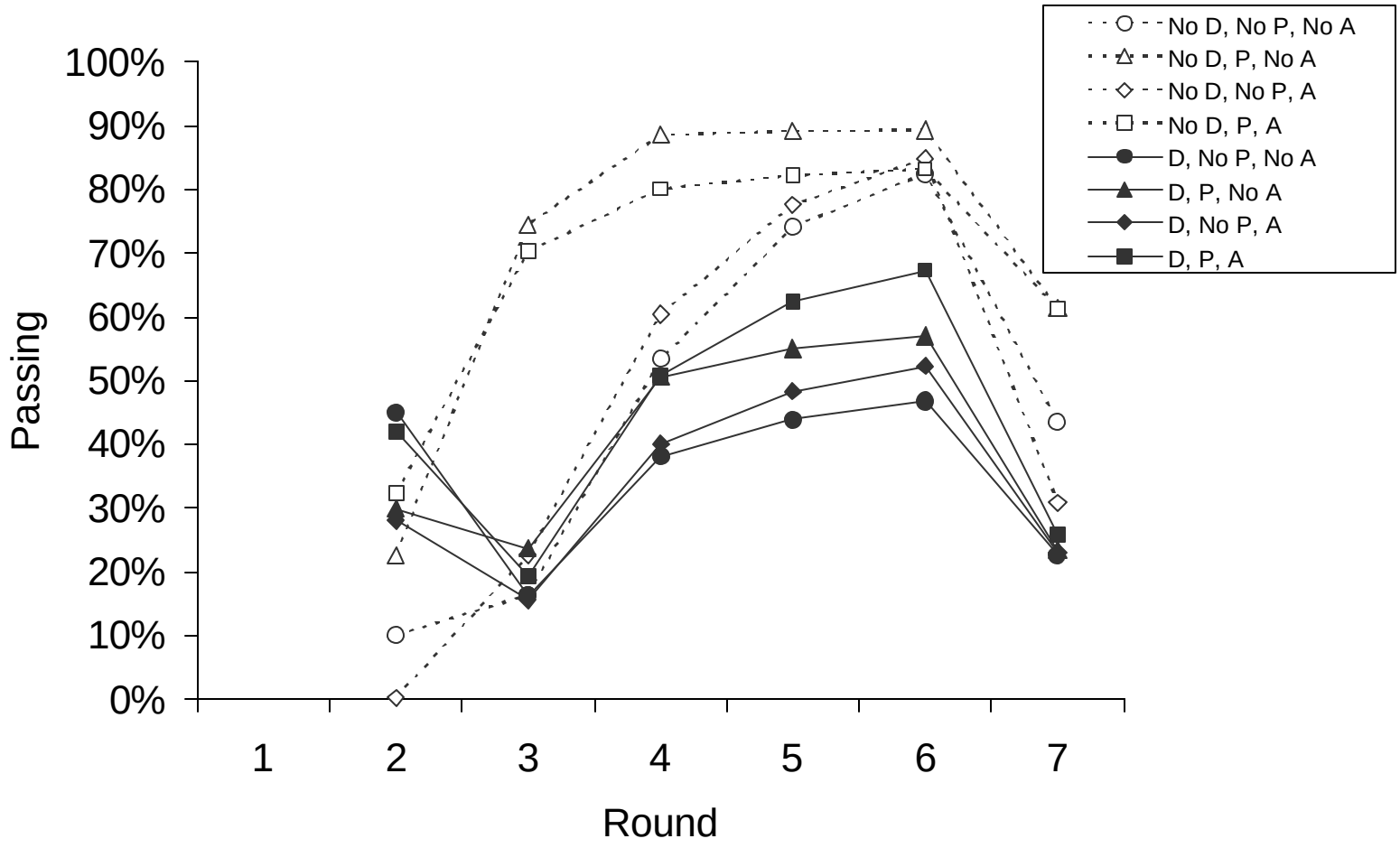
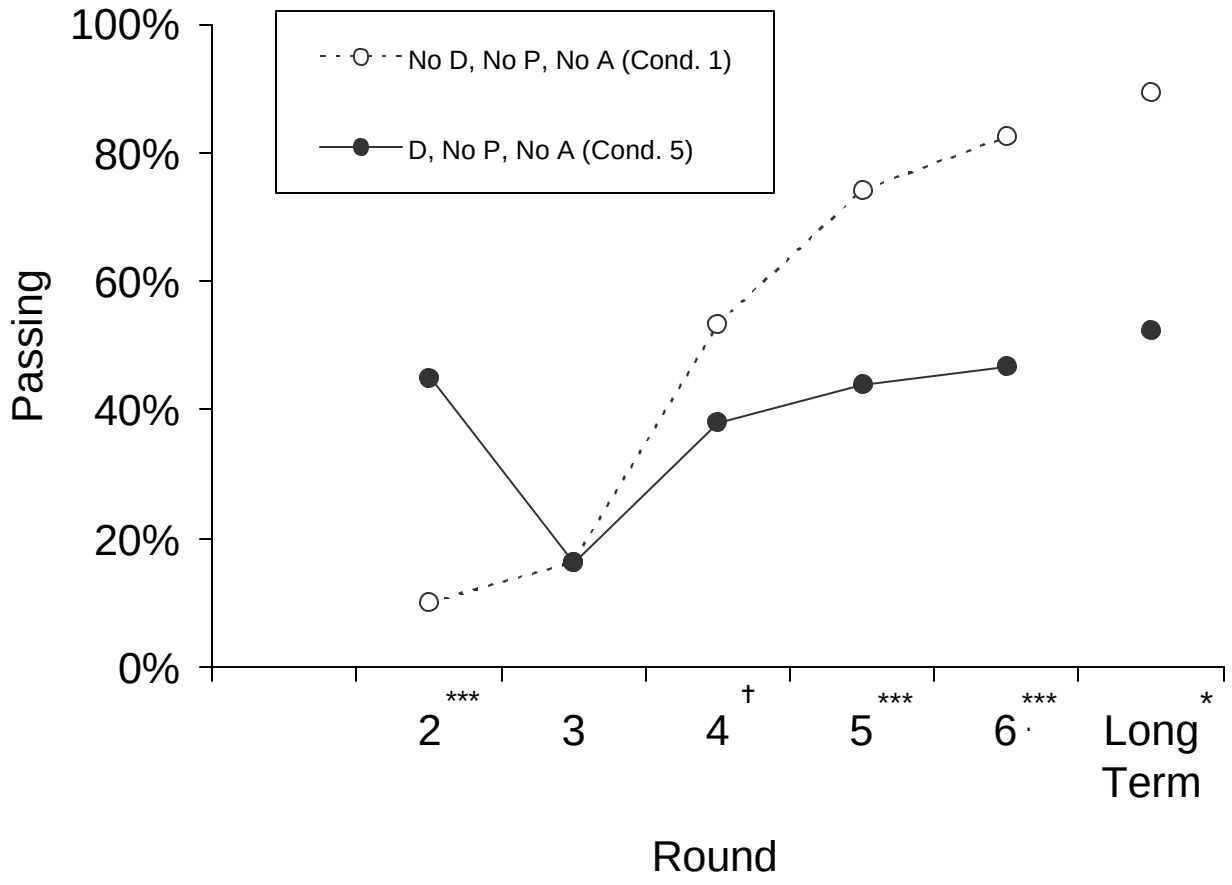


Figure 6: Deception and Trust Recovery: Fitted Values



For Figures 6 through 12, the significance of differences in each round is indicated by the following:
† p<.10, * p<.05, ** p<.01, *** p<.001

Figure 7: Promise and Trust Recovery: Fitted Values

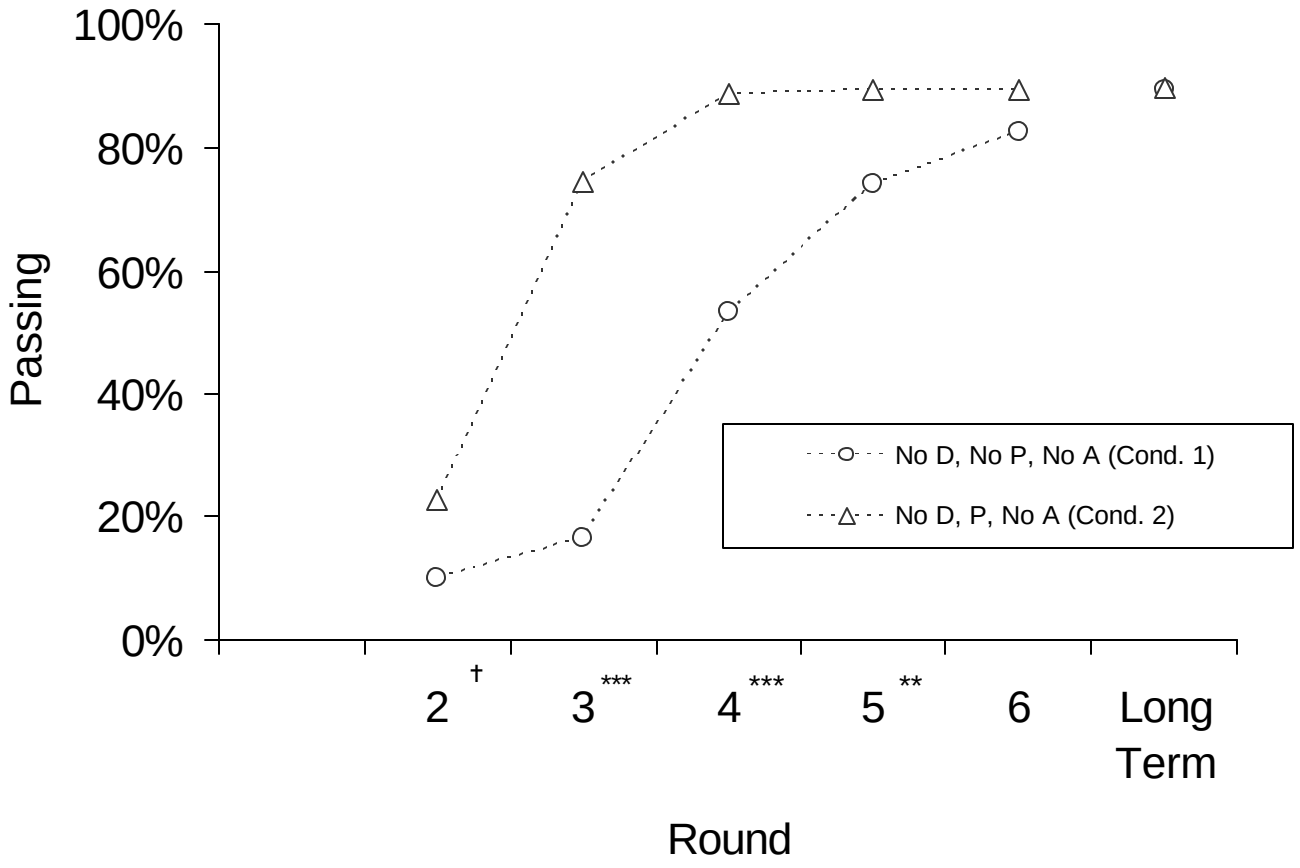


Figure 8: Apology and Trust Recovery: Fitted Values

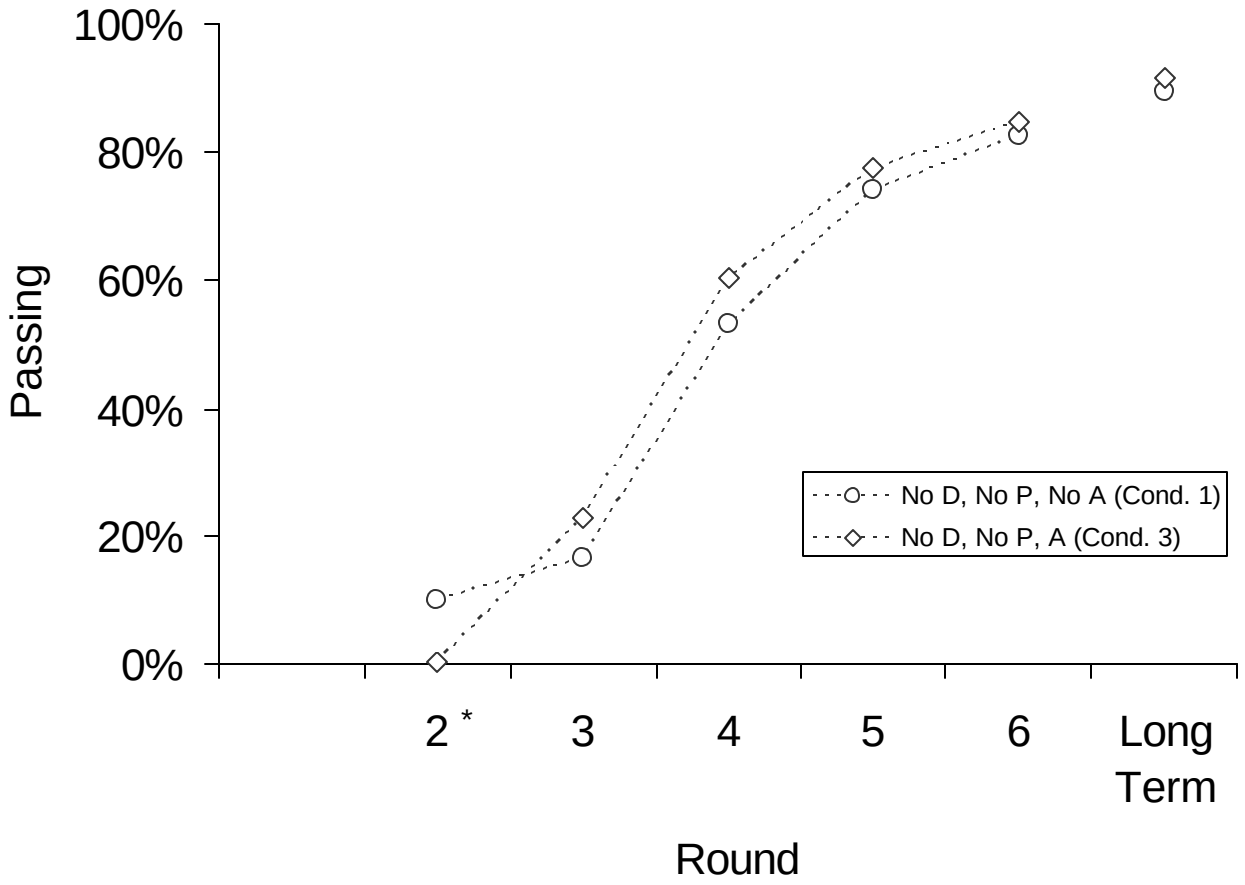


Figure 9: Promise and Apology and Trust Recovery: Fitted Values

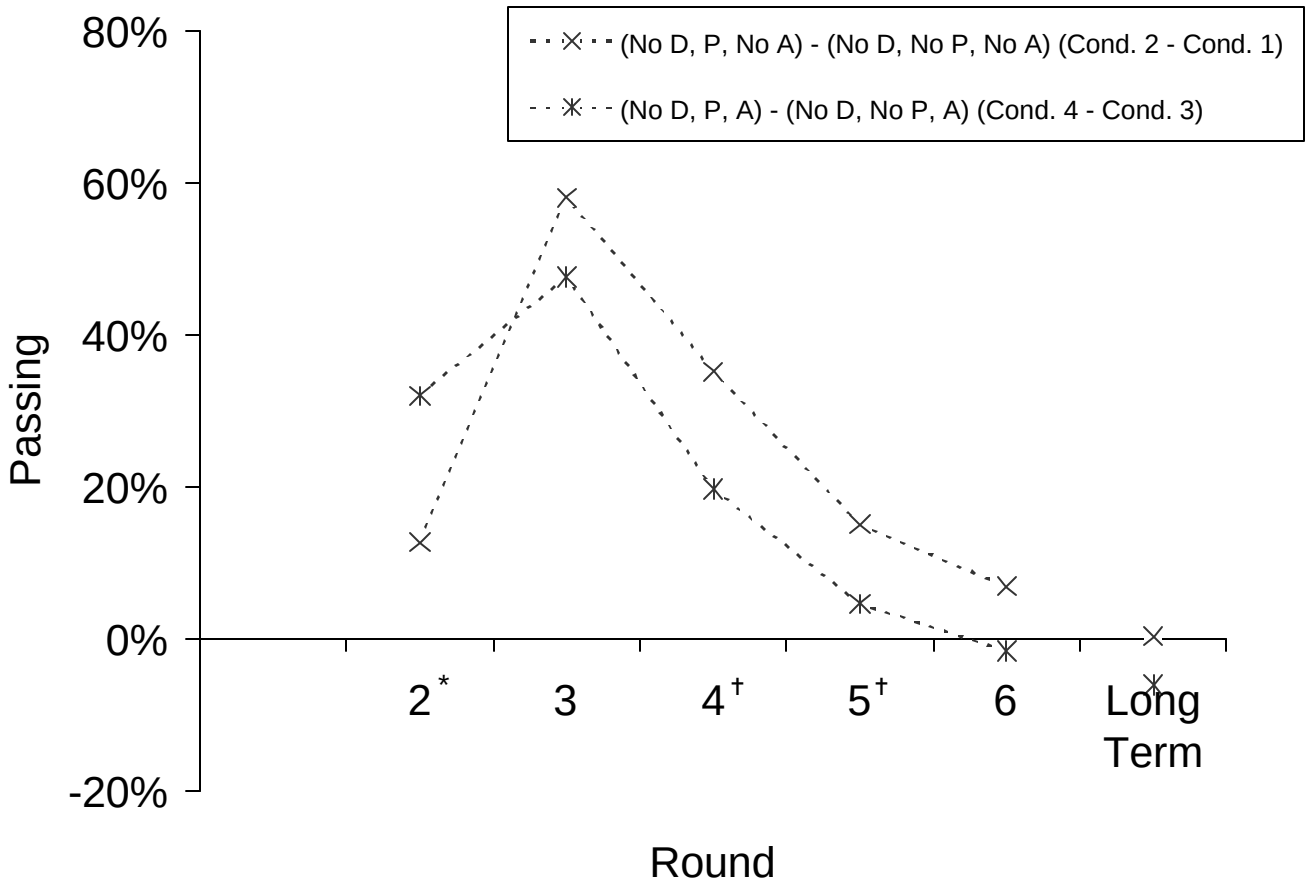


Figure 10: Deception and Promise and Trust Recovery: Fitted Values

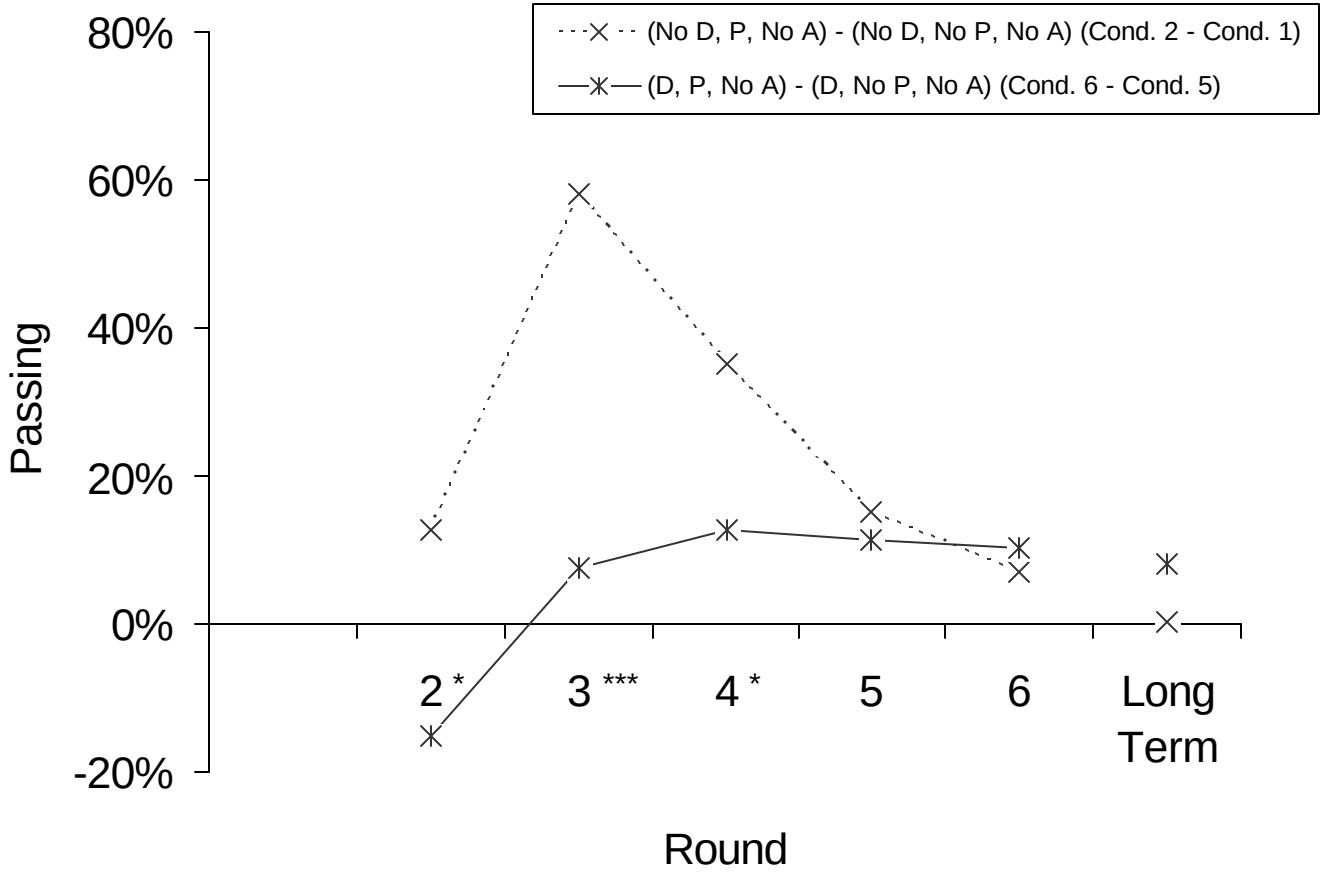


Figure 11: Deception and Apology and Trust Recovery: Fitted Values

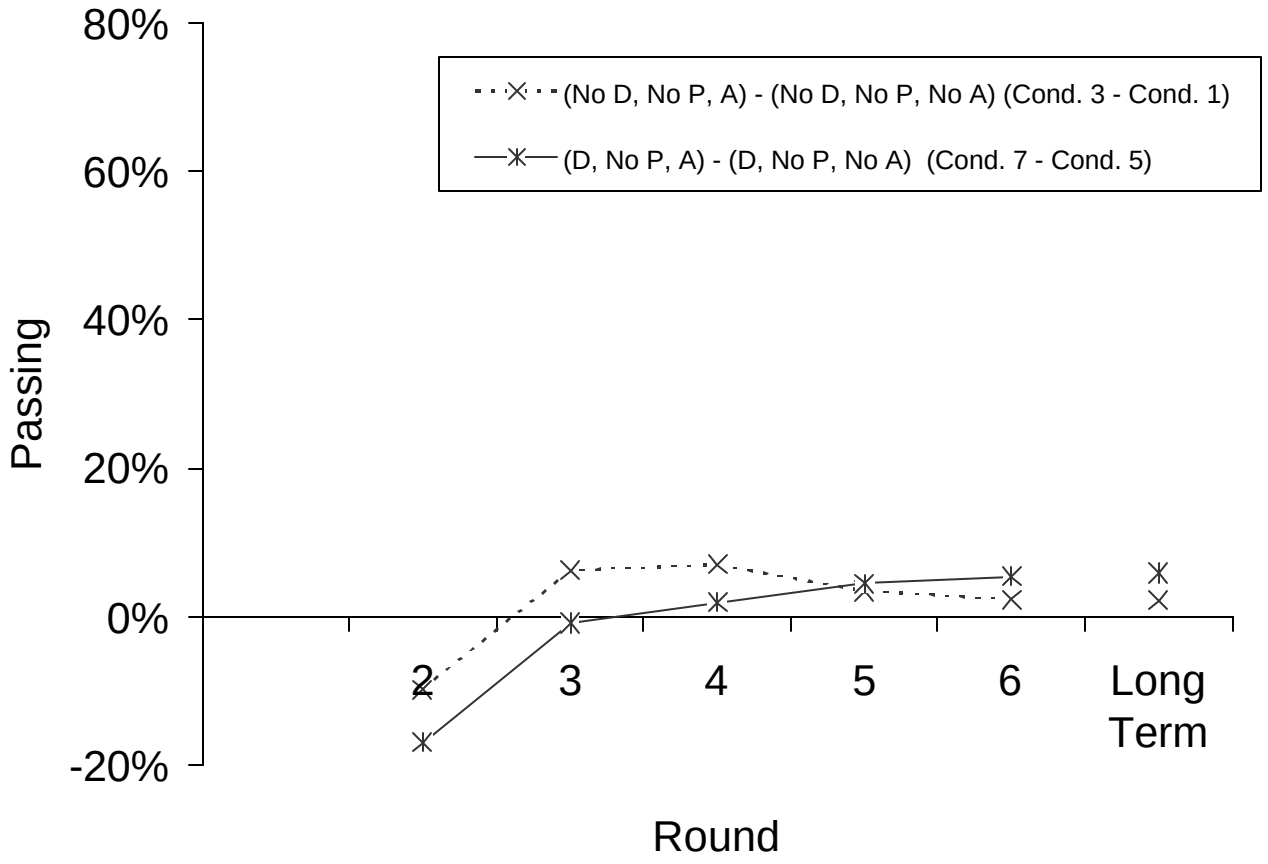


Figure 12: Deception, Promise, and Apology on Trust Recovery: Fitted Values