



University of Pennsylvania
ScholarlyCommons

Finance Papers

Wharton Faculty Research


1-2015

Prosocial Lies: When Deception Breeds Trust

Emma E. Levine
University of Pennsylvania

Maurice E. Schweitzer
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/fnce_papers

 Part of the [Finance and Financial Management Commons](#), and the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial Lies: When Deception Breeds Trust. *Organizational Behavior and Human Decision Processes*, 126 88-106. <http://dx.doi.org/10.1016/j.obhdp.2014.10.007>

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/fnce_papers/88
For more information, please contact repository@pobox.upenn.edu.

Prosocial Lies: When Deception Breeds Trust

Abstract

Philosophers, psychologists, and economists have long asserted that deception harms trust. We challenge this claim. Across four studies, we demonstrate that deception can increase trust. Specifically, prosocial lies increase the willingness to pass money in the trust game, a behavioral measure of benevolence-based trust. In Studies 1a and 1b, we find that altruistic lies increase trust when deception is directly experienced and when it is merely observed. In Study 2, we demonstrate that mutually beneficial lies also increase trust. In Study 3, we disentangle the effects of intentions and deception; intentions are far more important than deception for building benevolence-based trust. In Study 4, we examine how prosocial lies influence integrity-based trust. We introduce a new economic game, the *Rely-or-Verify* game, to measure integrity-based trust. Prosocial lies increase benevolence-based trust, but harm integrity-based trust. Our findings expand our understanding of deception and deepen our insight into the mechanics of trust.

Keywords

deception, trust, experiments, ethics

Disciplines

Business | Finance and Financial Management | Social and Behavioral Sciences

Prosocial Lies: When Deception Breeds Trust

Emma E. Levine

JMHH 533.6, 3730 Walnut Street
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104, United States
Email: emmased@wharton.upenn.edu

Maurice E. Schweitzer

JMHH 544, 3730 Walnut Street
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104, United States
Email: schweitz@wharton.upenn.edu

Cite as: Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, 126, 88-106.

Acknowledgements:

The authors thank Jessica Kennedy, Kristin Smith-Crowe, Keith Murnighan, and members of the OPIM-Decision Processes lab for their helpful feedback and advice. We thank the Wharton Behavioral Laboratory and the Wharton Risk Center Ackoff Doctoral Student Fellowship for financial support.

Running Head: DECEPTION BREEDS TRUST

Prosocial Lies: When Deception Breeds Trust

Abstract

Philosophers, psychologists, and economists have long asserted that deception harms trust. We challenge this claim. Across four studies, we demonstrate that prosocial deception can increase trust. Specifically, prosocial lies increase the willingness to pass money in the trust game, a behavioral measure of benevolence-based trust. In Studies 1a and 1b, we find that altruistic lies increase trust when deception is directly experienced and when it is merely observed. In Study 2, we demonstrate that mutually beneficial lies also increase trust. In Study 3, we disentangle the effects of intentions and deception; intentions are far more important than deception for building benevolence-based trust. In Study 4, we show that although prosocial lies increase benevolence-based trust, they harm integrity-based trust. To measure integrity-based trust behavior, we introduce a new economic game, the *Rely-or-Verify* game. Our findings expand our understanding of deception and deepen our insight into the mechanics of trust.

Keywords: deception, trust, experiments

Trust is essential to organizations and interpersonal relationships (e.g., Blau, 1964; Golembiewski & McConkie, 1975; Dirks & Ferrin, 2001; Lewicki, Tomlinson, & Gillespie, 2006; Rempel, Holmes, & Zanna, 1985; Valley, Moag, & Bazerman, 1998). Trust increases leadership effectiveness (Atwater, 1988; Bazerman, 1994; Dirks, 2000), improves the stability of economic and political exchange (Hosmer, 1995), reduces transaction costs (Granovetter, 1985), facilitates cooperation (Valley et al., 1998), and helps firms and individuals manage risk (Sheppard & Sherman, 1998). Golembiewski and McConkie (1975, p. 131) argued that, “There is no single variable which so thoroughly influences interpersonal and group behavior as does trust.”

Consistent with prior research, we define trust as, “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another” (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 395). A significant body of research has documented the negative effects of violating trust. For example, trust violations can harm cooperation and bargaining outcomes (Lount, Zhong, Sivanathan, & Murnighan, 2008; Croson, Boles, & Murnighan, 2003), lower organizational commitment (Robinson, 1996), provoke retaliation (Bies & Tripp, 2006), and, in more serious cases, trigger organizational-level failures (Gillespie & Dietz, 2009).

Although there are many ways to harm trust, existing research identifies one behavior as particularly toxic to trust: deception (e.g., Santoro & Paine, 1993; Boles, Croson, & Murnighan, 2000; Carr, 1968; O’Connor & Carnavale, 1997; Schweitzer & Croson, 1999; Schweitzer, Hershey, & Bradlow, 2006; Croson et al., 2003; Bok, 1978). Prior research suggests that deception is theoretically, philosophically, and empirically damaging to interpersonal trust. For example, philosopher Sir Francis Bacon argued that

dishonesty deprives, “people of two of the most principal instruments for interpersonal action—trust and belief” (from “On Truth”, cited in Tyler & Feldman, 2006). Empirical research has also demonstrated that deception harms relationships (Ford, King, & Hollender, 1988; Lewis & Saarni, 1993; Tyler & Feldman, 2006), elicits negative affect (Planalp, Rutherford, & Honeycutt, 1988), decreases liking, (Tyler, Feldman, & Reichert, 2006) and triggers retaliation (Boles et al., 2000; Croson et al., 2003). Furthermore, trust scholars have found that acts of deception cause enduring harm to trust. Though individuals can often repair trust following a violation (e.g., Kim, Ferrin, Cooper, & Dirks, 2004; Schweitzer et al., 2006), trust violations accompanied by deception irrevocably harm trust (Schweitzer et al., 2006).

We challenge the prevailing assumption that deception harms trust. We argue that most philosophers, psychologists, and economists, have confounded deceptive behavior with selfish intentions and outcomes. As a result, prior research that has documented the harmful effects of deception may really tell us more about the consequences of selfish behavior than deception per se.

We break new ground by demonstrating that some forms of deception increase trust. Across four experiments, we demonstrate that prosocial lying can increase behavioral and attitudinal measures of interpersonal trust. Consistent with prior work, we define deception as *the transmission of information that intentionally misleads others* (see Murnighan, 1991; Boles et al., 2000; Gino & Shea, 2012). We define prosocial deception as a type of deception. Prosocial lies involve the transmission of information that misleads *and benefits* a target (Levine & Schweitzer, 2014).

Our program of research expands our understanding of trust by highlighting the primacy of benevolence for building interpersonal trust. In our investigation, we explore distinct forms of trust and deception. We are the first to demonstrate that some common forms of deception can increase trust.

We report results from a series of experiments. In Studies 1, 2, and 3, participants experienced or observed deception and made trust game decisions. Across these studies, we find that prosocial lies increase trust. This is true when deception is directly experienced (Study 1a) and when it is merely observed (Study 1b). This pattern is also true when the prosocial lies are mutually beneficial and help both the target and the deceiver (Study 2).

In Studies 3a and 3b, we disentangle the effects of lying from the effects of prosocial and selfish intentions. When we control for intentions, we find that deception itself has no effect on trusting behavior. In other words, the decision to pass in the trust game reflects perceptions of benevolence, which is not undermined by deception. Prosocial intentions, regardless of whether they are associated with deception or honesty, significantly increase benevolence-based trust. In Study 3b, we demonstrate that our results do not simply reflect a negative reaction to selfish behavior. Instead, we find that prosocial deception increases trust compared to a neutral control condition.

In our final study, we explore how prosocial deception influences distinct types of trust. The trust game measures benevolence-based trust. We introduce a new economic game, the Rely-or-Verify game, to measure integrity-based trust. Although prosocial lying increases trust rooted in benevolence, it harms trust rooted in integrity. We demonstrate that the same action can have divergent effects on different dimensions of

trust.

Prosocial lying

Prosocial lying is a common feature of everyday communication. For example, an employee may tell a colleague that they delivered an excellent presentation when they did not, or thank a gift giver for a gift they would have rather not received.

As children, we learn to tell prosocial lies to be polite (Talwar, Murphy, & Lee, 2007; Broomfield, Robinson, & Robinson, 2002). Prosocial deception is also common in adult relationships (Tyler & Feldman, 2004). Adults lie in roughly 20% of their everyday social interactions (DePaulo & Bell, 1996), and most of these lies are prosocial (DePaulo & Kashy, 1998).

Individuals' endorsement of prosocial lies reflects the broader approval of unethical behaviors that help others. For example, individuals are more willing to cheat when cheating restores equity (Gino & Pierce, 2009, 2010a; Schweitzer & Gibson, 2008), helps disadvantaged others (Gino & Pierce, 2010b), and when the spoils of cheating are shared with others (Wiltermuth, 2011; Gino, Ayal, & Ariely, 2013). With respect to deception, prior experimental work has found that individuals are more willing to tell prosocial lies than selfish lies (Erat & Gneezy, 2012) and perceive them to be more ethical (Levine & Schweitzer, 2014).

Prosocial lying serves a number of interpersonal aims. While many prosocial lies are motivated by an altruistic desire to protect relational partners (e.g. DePaulo & Kashy, 1998) or provide interpersonal support (Brown & Levinson, 1987; Goffman, 1967), other lies have both prosocial and self-serving motives. For example, prosocial lying can be used to avoid conflict and facilitate uncomfortable social situations. When a wife asks her

husband if she looks fat in her dress, the husband may lie not only to protect his wife's feelings, but also to avoid conflict and a lengthy discussion about diet and exercise.

In the present research, we define altruistic lies as, "*false statements that are costly for the liar and are made with the intention of misleading and benefitting a target*" (Levine & Schweitzer, 2014: p. 108). We define mutually beneficial lies as *false statements that are beneficial for the liar and are made with the intention of misleading and benefitting the target*. We conceptualize altruistic and mutually beneficial lies as a subset of prosocial lies. Consistent with Bok (1978), we also distinguish between prosocial lies and white lies. White lies involve small stakes and can be prosocial or self-serving. Unlike white lies, prosocial lies can have large stakes. For example, some doctors misrepresent prognoses to give their patients comfort in their final weeks of life (e.g., Iezzoni, Rao, DesRoches, Vogeli, & Campbell, 2012).

Prosocial lies and trust

Prosocial lies are particularly relevant to the study of trust because they reflect a conflict between two central antecedents of trust: benevolence and integrity. Trust reflects an individual's expectation about another person's behavior. In contrast with research that conceptualizes trust as a belief about one's ability to carry out organizational duties or effectively perform a particular job (Kim et al., 2004; Kim, Dirks, Cooper, & Ferrin, 2006; Ferrin, Kim, Cooper, & Dirks, 2007), we investigate trust with respect to interpersonal relationships. Consistent with prior research (e.g. Lewicki & Bunker, 1995; Rousseau et al., 1998), we conceptualize interpersonal trust as the willingness to be vulnerable to exploitation within an interpersonal interaction.

Scholars have converged on three qualities of the trustee (the individual who is

trusted) that uniquely influence interpersonal trust: benevolence, ability, and integrity (Mayer, Davis, & Schoorman, 1995; Butler, 1991). Benevolence reflects the degree to which an individual has positive intentions or a desire to help the truster (Butler & Cantrell, 1984; Mayer et al., 1995). Ability reflects an individual's technical skills, competence, and expertise in a specific context (e.g., Mayer et al., 1995; Giffin, 1967; Sitkin & Roth, 1993). Integrity reflects an individual's ethicality and reputation for honesty (Mayer et al., 1995; Butler & Cantrell, 1984). In this work, we investigate the tension between benevolence and integrity.

We postulate that when benevolence and integrity conflict, benevolence will often be more important than integrity. We draw on prior trust, corruption, and person-perception research, to build our thesis.

Prior trust research suggests that benevolence may be particularly important for trust in interpersonal interactions. In dyadic relationships, trust hinges on concerns related to exploitation (Barney & Hansen, 1994; Lewicki & Bunker, 1995; Bhattacharya, Devinney & Pillutla, 1998), and perceptions of benevolence can allay these concerns. This idea is consistent with work that has documented the importance of motive attributions, beliefs about a counterpart's intentions, in interpersonal trust (e.g., Weber, Malhotra, & Murnighan, 2004; Malhotra & Murnighan, 2002; Pillutla, Malhotra, & Murnighan, 2003; Dunn, Ruedy, & Schweitzer, 2012; Lount & Pettit, 2012). Individuals who are perceived to have benevolent motives are perceived to be less likely to exploit a potential truster, and consequently, are more likely to be trusted. For example, lower status individuals are perceived to have benevolent motives and engender greater trust (Lount & Pettit, 2012).

The relationship between perceived integrity and interpersonal trust is less clear. One challenge is the subjective nature of integrity: “the trustee adheres to a set of principles that the truster finds acceptable” (Mayer et al., 1995, p. 719; Kim et al., 2004). Another challenge is the frequent confounding of benevolence and integrity (e.g. Kim et al., 2004; Schweitzer et al., 2006). Extant trust research has studied behaviors that violate ethical principles *and* cause harm to others, reflecting low integrity and low benevolence. For example, prior work has studied lies that exploit others for financial gain (Koning, Steinel, Beest, & van Dijk, 2011; Steinel & De Dreu, 2004; Schweitzer et al., 2006). These lies violate the principle of honesty and demonstrate selfishness. Not surprisingly, these lies harm trust. However, an individual may also lie to *benefit* a counterpart. This behavior violates the principle of honesty, but demonstrates benevolence. Existing research does not give us insight into how individuals might resolve these competing signals.

However, research on corruption and favoritism provides evidence that individuals, at times, place enormous trust in others who have demonstrated low integrity. For example, there is high trust within crime rings (Baccara & Bar-Isaac, 2008; Bowles & Gintis, 2004) and within communities that have been influenced by organized crime (Meier, Pierce, & Vaccaro, 2013). In these cases, there is high trust among in-group members, but low trust in out-group members. Individuals within the group are trusted because they care for and protect in-group members, even if they signal low integrity with respect to their relationships with out-group members.

We conjecture that the concern for benevolence is more deeply rooted than the concern for integrity in interpersonal relationships. A preference for ethical rules, such as

fairness and honesty, may be rooted in the broader concern of protecting people from harm (Gray, Young, & Wayts, 2012; Turiel, 1983). That is, benevolence may be the primary concern and integrity may be a derivative, secondary concern. Consistent with this proposition, Levine & Schweitzer (2014) found that when honesty harms other people and deception does not, honesty is perceived to be less ethical than deception.

We hypothesize that an individual will engender trust when they project high benevolence, even if they also project low integrity. Perceptions of benevolence, rather than perceptions of integrity, may explain why individuals are willing to be vulnerable to interpersonal exploitation. As a result, we hypothesize that prosocial lies, which demonstrate high benevolence, but low integrity, can build trust.

In our first three studies, we focus our investigation on the positive relationship between prosocial lies and trust. In these studies, we explore the relationship between prosocial lying and passing in the trust game, a benevolence-based measure of trust. In our final study, we explore the boundaries of the relationship between deception and trust by examining the relationship between prosocial lies and integrity-based trust.

Overview of current research

Across our studies, we use deception games (adapted from Erat & Gneezy, 2012; Cohen, Gunia, Kim-Jun, & Murnighan, 2009; Gneezy, 2005) and trust games (adapted from Berg, Dickhaut, & McCabe, 1995). We use deception games to operationalize prosocial lies because these games allow us to cleanly manipulate the intentions associated with deception and, consequently, draw causal inferences about the role of intentions and deception in building trust.

We use the trust game because it operationalizes the fundamental components of an interpersonal trusting decision: the willingness to be vulnerable based on positive expectations of another (Rousseau et al., 1998; Pillutla et al., 2003). The trust game is also the predominant paradigm used to measure trust throughout psychology and economics (e.g., Berg et al., 1995; Schweitzer et al., 2006; McCabe, Rigdon, & Smith, 2003; Glaeser, Laibson, Scheinkman, & Soutter, 2000; Malhotra & Murnighan, 2002; Malhotra, 2004). In the standard trust game, the truster is endowed with money and has the opportunity to keep the money or pass the money to the trustee. The amount of money grows if the truster passes it to the trustee. The trustee then has the opportunity to either return some portion of the money to the truster or keep all of the money for himself. The truster's initial decision to pass money represents trust (Pillutla et al., 2003; McCabe et al., 2003; Glaeser et al., 2000; Malhotra & Murnighan, 2002; Malhotra, 2004). Though trust game decisions may also reflect preferences for equality and risk (Ashraf, Bohnet, & Piankov, 2003), the external validity of trust game decisions has been documented with financial investment decisions (e.g., Karlan, 2005) and prior work that has closely linked trust game behavior with attitudinal measures of trust (e.g., Houser, Schunk, & Winter, 2010; Schweitzer et al., 2006).

Across our studies, we investigate different types of prosocial lies. We begin by examining the consequences of altruistic lies. In Study 1a, participants were paired with a confederate who either told an altruistic lie to the participant or was selfishly honest. Participants then played a trust game with the confederate. In Study 1a, we find that being deceived increases trust; participants were more trusting of confederates who lied to them than confederates who were honest. In Study 1b, we rule out reciprocity as an

alternative explanation. In this study, participants observed, rather than experienced, altruistic deception and then made trust decisions. We find that individuals trust altruistic liars, even when they did not benefit from the prosocial deception.

In Study 2, we extend our investigation by examining different types of lies. In this study, we find that even when prosocial lying helps the liar, deception increases trust; non-altruistic prosocial lies, and mutually beneficial lies increase trust. In Studies 3a and 3b, we isolate the effects of intentions and deception by manipulating them orthogonally. In Study 3a, we find that deception itself has no direct effect on interpersonal trust, but that intentions matter immensely. Prosocial individuals who told lies or were honest were trusted far more than selfish individuals who lied or were honest. In Study 3b, we include two control conditions and demonstrate that relative to control conditions, prosocial intentions increase trust and selfish intentions decrease trust.

In our final study, we explore how prosocial lying influences distinct types of trust. We introduce a new economic game, the Rely-or-Verify game, to measure integrity-based trust. We manipulate prosocial lying using a repeated deception game and find that prosocial lying increases benevolence-based trust, but decreases integrity-based trust. Across all of our studies, our sample size or the number of days that the study would run was determined in advance, and no conditions or variables were dropped from any analyses we report (Simmons, Nelson, & Simonsohn, 2011).

Study 1

In Study 1a and 1b, we explore the relationship between altruistic lying and trusting behavior. In Study 1a, participants played a trust game with a counterpart who either lied to them or told them the truth. In Study 1b, participants observed an individual

who told an altruistic lie or a selfish truth to a third party. Together, Studies 1a and 1b demonstrate that altruistic deception can increase trust and that this result is not driven by direct reciprocity.

Study 1a

Method

Participants. We recruited 125 adults to participate in an online study in exchange for payment via Amazon Mechanical Turk.

Procedure and Materials. In this study, we randomly assigned participants to one of two conditions in a between-subjects design. Participants played a deception game with an individual who either told an altruistic lie or was selfishly honest. Participants then played a trust game with the same partner.

Manipulation of altruistic lies. We used a modified deception game (Erat & Gneezy; 2012; Cohen et al., 2009; Gneezy, 2005; Levine & Schweitzer, 2014) to operationalize altruistic lies. We referred to the deception game as “Exercise 1” in the experiment.

In our version of the deception game, two individuals were paired and randomly assigned to the role of either Sender or Receiver. The payoffs for each pair of participants (one Sender and one Receiver) were determined by the outcome of a computer-simulated coin flip and the choices the participants made. In the deception game, the Sender had the opportunity to lie to the Receiver about the outcome of the coin flip. In the experiment, we refer to the potential liar as “the Sender.”

The deception game unfolded in the following steps:

1. Senders were told the outcome of a computer-simulated coin flip. In our study, the coin always landed on heads.
2. The Sender then had to report the outcome of the coin flip to his/her partner, the Receiver. The Sender could send one of two possible messages to the Receiver. The message could read, “The coin landed on heads” or “The coin landed on tails.”
 - The Sender knew that the outcome the Receiver chose (heads or tails) determined the payment in the experiment. The Sender also knew that the only information the Receiver would have was the message from the Sender and that most Receivers chose the outcome indicated in the Sender’s message.
 - The Sender knew there were two possible payment options, A and B. If the Receiver chose the correct outcome, the Sender and the Receiver would be paid according to Option A. Otherwise, the Sender and the Receiver would be paid according to Option B.
3. In Study 1a, Option A was \$2 for the Sender and \$0 for the Receiver. Option B was \$1.75 for the Sender and \$1 for the Receiver. Throughout our studies, we manipulated the payments associated with Option A and Option B to operationalize different types of lies. We summarize the payoffs associated with each choice in Table 1.
4. After receiving the Sender’s message, the Receiver had to choose an outcome: heads or tails. The Receiver knew that his/her choice determined the payment in the experiment, but the Receiver did not know the payoffs

associated with the choice. The Sender's message was the only piece of information the Receiver had.

Therefore, Senders faced the following options:

- A. Send an honest message, e.g. "*The coin landed on heads.*"

Honesty was most likely to lead to an outcome that was costly to the Receiver, and benefitted the Sender (i.e. selfish).

- B. Send a dishonest message, e.g. "*The coin landed on tails.*"

Lying was most likely to lead to an outcome that benefitted the Receiver, and was costly to the Sender (i.e. altruistic).

In Study 1a, we assigned all participants to the role of Receiver and informed them that their decisions would be matched with the decisions of a previous participant, who had been assigned to the role of Sender. After reading the instructions for the deception game and passing a comprehension check¹, participants received a message from their partner, the Sender. The Sender's message either read "The coin landed on heads" (the *Selfish Honesty* condition) or "The coin landed on tails" (the *Altruistic Lie* condition). Participants then made their prediction by choosing either "Heads" or "Tails."² Participants did not know the possible payoffs when they made their choice.

After making their choice, participants learned more information about the deception game. Specifically, we gave them all of the Sender's private information.

¹ Participants had to pass two comprehension checks, one for the deception game and one for the trust game, in order to complete the entire study. Participants who failed a comprehension check had the opportunity to reread the instructions for the exercise and retake the comprehension check. If any participant failed a comprehension check twice, they were not allowed to complete the study. We followed this procedure in every study.

² A total of 89% of participants actually chose the outcome indicated in their partner's message. Whether or not participants chose the outcome indicated in the message did not influence our results. That is, our results are not influenced by whether or not participants were successfully deceived.

Participants learned that the Sender knew the coin had landed on heads. Therefore, participants learned that the Sender either lied to them or had been honest. In addition, participants learned the payoffs associated with the Sender's choice. Therefore, participants learned that lying was altruistic and honesty was selfish. This was our manipulation of altruistic lying.

After participants learned about the information their partner knew as the Sender in the deception game, participants played a trust game with the Sender. We referred to the trust game as "Exercise 2" in the experiment. We ran a pilot study with a non-overlapping sample ($N = 40$) in order to generate real decisions with which to match the decisions of participants in our main study.

The trust game. In our trust game, we assigned all participants to the role of Player 1 and told them that they would be paired with the Sender with whom they had just been paired with in Exercise 1 (the deception game). In our version of the trust game, Player 1 was given \$1 and could make one of two choices: "Keep \$1" or "Pass \$1." Choosing "Keep \$1" led to a payout of \$1 for Player 1 and \$0 for his partner, Player 2. If Player 1 passed the \$1, the money would grow to \$3 and Player 2 could then either choose to "Keep \$3" or "Return \$1.50."

Dependent variables.

Trusting behavior. The choice to pass money in the trust game served as our primary dependent variable. In addition, after making a decision, participants rated their partner's trustworthiness, benevolence, and deception. For all attitudinal measures, we used 7-point Likert scales anchored at 1 = "Strongly disagree" and 7 = "Strongly agree".

Attitudinal trust. We measured attitudinal trust with two items ($r(121) = .89$): “I trust Player 2 to RETURN money,” and “I am confident that Player 2 will RETURN money.”

Perceived benevolence. We measured perceived benevolence using three items ($\alpha = .80$): “This person is [kind, nice, and selfish (reverse-scored)].”

Perceived deception. We used three items to measure the extent to which our manipulation was recognized as deception ($\alpha = .84$): “This person sent an honest message as a Sender in Exercise 1” (reverse-scored), “This person lied about the outcome of the coin flip in Exercise 1,” and, “This person deceived his/her partner in Exercise 1.”

After participants submitted their responses, we asked two multiple-choice recall questions,³ collected demographic information, and asked participants what they thought the purpose of the study was. Participants then received a bonus payment based upon their decisions.

Results

We report results from the 121 adults (45% female; $M_{\text{age}} = 32$ years, $SD = 9.77$) who passed all comprehension checks and completed the entire study; 4 participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Trusting behavior. Supporting our thesis, participants were significantly more likely to trust a partner who told them an altruistic lie (56%), than a partner who was honest (32%), $\chi^2(1, N = 121) = 6.88, p < .01$. Figure 1 depicts these results.

³ In every study, at least 80% of participants were able to recall the manipulation at the end of the study. For each study, we report analyses for the entire sample, but our results are unchanged when we restrict our sample to only those who answered the recall questions correctly.

Attitudinal Trust. Our attitudinal trust measure parallels our behavioral trust results. Participants reported that they trusted their partners more in the *Altruistic Lie* condition ($M = 3.77, SD = 1.91$) than in the *Selfish Honesty* condition ($M = 2.72, SD = 1.76$), $F(1, 119) = 9.85, p < .01$. Our behavioral and attitudinal measures of trust were highly correlated, $r(121) = .89, p < .001$, suggesting that passing decisions reflected trust beliefs.

Perceived Benevolence. Participants also perceived their partners to be more benevolent in the *Altruistic Lie* condition ($M = 4.19, SD = 1.55$) than in the *Selfish Honesty* condition ($M = 3.45, SD = 1.32$), $F(1, 119) = 8.12, p < .01$.

Perceived Deception. Consistent with our manipulation, participants also perceived their partners to be more deceptive in the *Altruistic Lie* condition ($M = 5.37, SD = 1.35$) than in the *Selfish Honesty* condition ($M = 2.88, SD = 1.34$), $F(1, 119) = 102.60, p < .001$.

Discussion

Consistent with our thesis, individuals trusted altruistic liars more than honest partners. Importantly, participants recognized that they had been deceived, but rated their counterparts as more benevolent and thus, more trustworthy. Study 1a provides initial evidence that deception can increase trust.

Figure 1 about here

Study 1b

In Study 1a, participants who were deceived directly benefitted from the deception. Their subsequent trust decisions may have been influenced by reciprocity. In Study 1b, we rule out reciprocity as an alternative explanation. In Study 1b, participants observe, rather than experience, deception. Individuals played a trust game with counterparts who either had or had not told an altruistic lie to a different partner in a previous interaction.

Method

Participants. We recruited 261 participants from a city in the northeastern United States to participate in a study in exchange for a \$10 show-up fee.

Procedure and Materials. In this study, we randomly assigned participants to one of two conditions in a between-subjects design. Participants observed an individual who either told a prosocial lie or was selfishly honest and then played a trust game with this person.

We seated participants in separate cubicles to complete this study on the computer. The study was titled, “Partner Exercises.” We told participants that they would complete two separate exercises with two separate partners. The first exercise, which we called “Exercise 1,” was a deception game. Within the experiment, we called the second exercise, the trust game, “Exercise 2.” Both games are similar to the games we used in Study 1a. In Study 1b, however, we matched participants with two different partners. Participants first completed the deception game and chose Heads or Tails. We paired participants with a new partner for the trust game. Participants did not learn about their own outcome in the deception game until they completed the entire study.

Manipulation of altruistic lies. We told participants that their partner in the trust game (“Exercise 2”) had been matched with a different participant in the deception game (“Exercise 1”) and had been assigned to the role of Sender. We then revealed the decision the Sender had made and the information they had had prior to making that decision. As in Study 1a, by revealing the Sender’s decision and the payments associated with their choice, participants learned that the Sender either told an altruistic lie or was selfishly honest.

The trust game. The trust game in Study 1b was similar to the trust game we used in Study 1a. We assigned every participant to the role of Player 1 and we matched each participant with a Player 2 who was the Sender in the first Exercise. In the trust game in Study 1b, participants started with \$2. If Player 1 chose to “Pass \$2” the money grew to \$5. If Player 1 passed the money, Player 2 had the decision to either “Keep \$5” or “Return \$2.50.” We used larger stakes in this study than those we used in Study 1a because our participants were university students, rather than Mechanical Turk participants.

Dependent variables. As in Study 1a, our main dependent variable was trusting behavior, measured by the decision to pass money in the trust game. All of our other dependent variables were identical to those we collected in Study 1a ($r > .87$; α 's $> .80$).

After participants submitted their responses, we asked two multiple-choice recall questions, collected demographic information, and asked participants what they thought the purpose of the study was. Participants then received bonus payment based on their decisions.

Results

We report the results from 257 participants (60.3% female; $M_{\text{age}} = 20$ years, $SD = 2.30$) who passed all comprehension checks and completed the entire study; 4 participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Trusting behavior. Consistent with our prediction, participants were more likely to trust their partner when they learned that their partner had told someone else an altruistic lie (39%), than when they learned that their partner had told someone else the truth (21%), $\chi^2(1, N = 257) = 9.79, p < .01$. We depict these results in Figure 1.

Attitudinal Trust. As in Study 1a, our behavioral and attitudinal measures of trust followed the same pattern and were highly correlated, $r(257) = .70, p < .001$. Participants reported trusting their partners more in the *Altruistic Lie* condition ($M = 3.51, SD = 1.71$) than in the *Selfish Honesty* condition ($M = 2.66, SD = 1.46$), $F(1, 255) = 18.04, p < .01$.

Perceived Benevolence. Participants also perceived their partners to be more benevolent in the *Altruistic Lie* condition ($M = 4.14, SD = 1.39$) than in the *Selfish Honesty* condition ($M = 3.75, SD = 1.07$), $F(1, 255) = 8.12, p = .01$.

Perceived Deception. Consistent with our manipulation, participants also perceived their partners to be more deceptive in the *Altruistic Lie* condition ($M = 4.91, SD = 1.45$) than in the *Selfish Honesty* condition ($M = 3.40, SD = 1.64$), $F(1, 255) = 60.18, p < .001$.

Discussion

As in Study 1a, our participants trusted altruistic liars more than people who were selfishly honest. In this study, participants observed rather than experienced deception. Results from this study rule out direct reciprocity as an alternative explanation for our findings in Study 1a. Unlike Study 1a, participants in this study did not benefit from the act of deception.

Study 2

In Study 2, we extend our investigation by examining how different types of prosocial lies influence trust. In Studies 1a and 1b, we investigated altruistic lies. Because these lies were costly for the liar, it is possible that our trust results are driven in part by the desire to compensate the liar for his altruism. We rule out this explanation in Study 2.

In Study 2, we demonstrate that our findings extend to prosocial lies that are not characterized by altruism. We explore how non-altruistic prosocial lies, lies that help the deceived party and have no effect on the liar, and mutually beneficial lies, lies that benefit the deceived party *and* the liar, influence trust.

Method

Participants. We recruited 300 adults to participate in an online study in exchange for payment via Amazon Mechanical Turk.

Procedure and Materials. As in Study 1b, participants learned about the decisions an individual made as a Sender in a deception game and then played a trust game with that individual. In this study, we randomly assigned participants to one of four experimental conditions in a 2(Deception: Lie vs. Honesty) x 2(Type of lie: Prosocial vs. Mutually beneficial) between-subjects design. That is, participants learned the following about a Sender in the deception game: the Sender either lied or was honest; and lying

either had no effect on the Sender and benefited the Receiver (i.e. was prosocial) or benefited both the Sender and the Receiver (i.e. was mutually beneficial).

In this study, participants first learned that they would play a trust game with a partner. We referred to the trust game as “The Choice Game” in the experiment. After participants learned about the trust game, but before they made any decisions, we told them that they would learn more information about their partner. Participants learned that their partner in the trust game had completed the trust game, along with another exercise, “The Coin Flip Game,” in a previous study. “The Coin Flip Game” was the same deception game as the one we used in Studies 1a and 1b. Participants in this study, however, observed but did not play the deception game. That is, our participants did not have a chance to earn money before they played the trust game.

Manipulation of prosocial lies. We told participants that their partner in the trust game had been matched with a different participant in the deception game (“The Coin Flip Game”) and had been randomly assigned to the role of Sender. We then explained the deception game and revealed the Sender’s decision in that game.

In Study 2, we manipulated both the decision to lie and the type of lie that was told. In order to manipulate the type of lie, we manipulated the payments associated with Outcome A (Honesty) and Outcome B (Lying). When lying was prosocial, Outcome A yielded \$2 for the Sender, \$0 for the Receiver and Outcome B yielded \$2 for the Sender, \$1 for the Receiver. That is, this lie was prosocial, but not altruistic. When lying was mutually beneficial, Outcome A yielded \$2 for the Sender, \$0 for the Receiver and Outcome B yielded \$2.25 for the Sender, \$1 for the Receiver. We summarize the payments associated with each type of lie in Table 1.

Participants learned whether the Sender had been honest or had lied in the deception game, and whether or not lying was prosocial or mutually beneficial. Then, participants played the trust game with the Sender and rated the Sender.

The trust Game. We referred to the trust game as “The Choice Game” in the experiment. The trust game we used in Study 2 was similar to the one we used in Study 1a and Study 1b. In this version of the trust game, however, participants played with lottery tickets rather than monetary outcomes. Using lottery tickets allowed us to increase the stakes on Mechanical Turk (a chance to win \$25) and prevented participants from directly comparing outcomes in the deception game and the trust game.

In this trust game, we assigned participants to the role of Player 1 and matched them with the confederate Player 2 who had made decisions in “The Coin Flip Game.” In the trust game, Player 1 started with 4 lottery tickets. If Player 1 chose to “Keep 4 lottery tickets,” Player 1 earned 4 lottery tickets and Player 2 earned 0 lottery tickets. If Player 1 chose to “Pass 4 lottery tickets,” the number of tickets tripled to 12 tickets and Player 2 made the decision to either “Keep 12 lottery tickets” or “Return 6 lottery tickets.” Participants knew that the more tickets they had, the more likely they were to win the \$25 lottery at the end of the study.

Dependent variables. Our main dependent variable was trusting behavior, measured by Player 1’s decision to pass the lottery tickets in the trust game. Our measures of trusting attitudes and perceived deception were identical to those we collected in Studies 1a and 1b ($r > .93$; α ’s $> .82$). We modified our measure of perceived benevolence to include new items that were more specific: “This person is benevolent”, “This person would not purposefully hurt others”, “This person has good intentions” ($\alpha =$

.86). We used a 7-point Likert scale anchored at 1 = “Strongly disagree” and 7 = “Strongly agree.”

After participants submitted their responses, we asked two multiple choice recall questions, collected demographic information, and asked participants what they thought the purpose of the study was. We then told participants the number of lottery tickets they received as a result of their decision and their counterpart’s decision in the trust game. We conducted the lottery the day the experiment ended.

Results

We report the results from 293 participants (39.9% female; $M_{\text{age}} = 32$ years, $SD = 11.2$) who passed the comprehension checks and completed the entire study; 7 participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Trusting behavior. We first conducted a logistic regression on trusting behavior, using *Deception*, *Type of Lie*, and the *Deception x Type of Lie* interaction as independent variables. We found a main effect of *Deception* ($b = .557$, $p < .01$), such that participants were more trusting of individuals who told lies that helped others. Specifically, 63% of participants trusted partners who had lied, whereas only 37% of participants trusted partners who had been honest; $\chi^2(1, N = 293) = 20.23$, $p < .01$.

We found no main effect of *Type of Lie* and we found no significant *Deception x Type of Lie* interaction ($ps > .32$). Although lying had a directionally larger effect on trust when the prosocial lie was not mutually beneficial, this difference was not significant. In

Figure 2, we display the percentage of participants who passed money in each of our four experimental conditions.

Figure 2 about here

Attitudinal Trust. As in Studies 1a and 1b, our behavioral and attitudinal measures of trust were highly correlated, $r(293) = .73, p < .001$ and follow the same pattern. A two-way ANOVA revealed a main effect of *Deception* on attitudinal trust, $F(1,289) = 16.42, p < .001$. Participants perceived their partner to be more trustworthy when they lied ($M = 3.83, SD = 1.88$) than when they had told the truth ($M = 2.95, SD = 1.91$). We do not find a main effect of *Type of Lie*, $F(1,289) = .13, p = .71$, nor do we find a significant *Deception* x *Type of Lie* interaction, $F(1,289) = .34, p = .56$.

Perceived Benevolence. A two-way ANOVA also revealed a main effect of *Deception* on perceived benevolence, $F(1,289) = 16.42, p < .001$. Participants perceived their partner to be more benevolent when they lied ($M = 4.56, SD = 1.15$) than when they told the truth ($M = 3.63, SD = 1.33$).

We also found a marginally significant *Deception* x *Type of Lie* interaction, $F(1,289) = 3.28, p = .07$. Lying had a greater effect on perceived benevolence when the lie was prosocial ($M_{\text{lie}} = 4.73, SD_{\text{lie}} = 1.22$ vs. $M_{\text{honesty}} = 3.51, SD_{\text{honesty}} = 1.37$), $t(138) = 5.77, p < .001$; than when the lie was mutually beneficial ($M_{\text{lie}} = 4.44, SD_{\text{lie}} = 1.08$ vs. $M_{\text{honesty}} = 3.75, SD_{\text{honesty}} = 1.29$), $t(153) = 3.43, p < .001$. We do not find a main effect of *Type of Lie*, $F(1,289) = .03, p = .86$.

Perceived Deception. Consistent with our manipulation, participants perceived their partners to be more deceptive when their partner had lied ($M = 5.39$, $SD = 1.24$) than when they told the truth ($M = 2.83$, $SD = 1.45$), $F(1,289) = 259.69$, $p < .001$. We do not find a main effect of *Type of Lie*, $F(1,289) = .01$, $p = .91$, nor do we find a significant *Deception x Type of Lie* interaction, $F(1,289) = 1.29$, $p = .26$.

Discussion

In Study 2, we demonstrate that altruism is not a necessary condition for lies to increase trust. Prosocial lies that are not costly for the liar and prosocial lies that benefit the liar both increase trust. These results suggest that trusting behavior does not simply reflect a desire to compensate a liar for altruism. Rather, individuals trust people who help others, even when that help involves deception *and* when that help is self-serving.

Although mutually beneficial lies are a weaker signal of benevolence than prosocial lies that do not benefit the deceiver, the self-serving nature of these lies did not undermine trust. These results suggest that for trust, judgments of benevolence may be more important than selflessness.

Study 3

Our initial studies demonstrate that prosocial lies can increase trust. In Studies 3a and 3b, we extend our investigation by independently manipulating deception and intentions (Study 3a) and by including two control conditions to disentangle the effects of selfishness from prosociality (Study 3b).

Study 3a

Method

Participants. We recruited 337 participants from a city in the northeastern United States to participate in a study in exchange for a \$10 show-up fee.

Procedure and Materials. We seated participants in separate cubicles to complete the study on the computer. The study was titled, “Partner Exercise.” As in Study 2, participants learned about the decision a Sender made in a deception game and then played a trust game with that Sender. In Study 3a, we randomly assigned participants to one of four experimental conditions in a 2(Deception: Lie vs. Honesty) x 2(Intentions: Altruistic vs. Selfish) between-subjects design. Specifically, participants observed a Sender who either lied or sent an honest message in a deception game, and whose choice was either altruistic or selfish. Participants then played a trust game with this partner.

Manipulation of lies. The deception game in Study 3a was similar to the one we used in our prior studies. In this game, however, we used a random number generator rather than a coin flip to begin the game. The game was otherwise identical to the game we used in Study 2. That is, the payoffs for each pair of participants (one Sender and one Receiver) were determined by the outcome of a random number generator and the choices made by the Sender and the Receiver. Senders knew the correct number was 4, and could send an honest message (e.g., “The number is 4”) or a dishonest message (e.g., “The number is 5”). We used a random number generator rather than a coin flip so that participants would be less likely to make strategic inferences about the message the Sender sent (e.g., The Sender sent the message: “The coin landed on heads”, hoping their partner would pick “tails”).

Importantly, Senders in this experiment played The Number Game with one of two possible payment structures. These payment structures enabled us to manipulate whether deception or honesty was associated with selfish or altruistic intentions.

The first payment structure was identical to the one we used in Studies 1a and 1b. This payment structure represented the choice between selfish honesty (Option A) and altruistic lying (Option B). In the second payment structure, we reversed the payoffs. This payment structure represented the choice between altruistic honesty and selfish lying.

After learning about the Sender's choice in the deception game, participants played a trust game with the Sender. We ran a pilot study with a non-overlapping sample ($N=41$) to generate decisions with which to match the decisions participants made in Study 3a.

The Trust Game. We referred to the trust game as “The Choice Game” in the experiment. “The Choice Game” was identical to the trust game we used in Study 1b. Participants had the choice to either “Keep \$2” or trust their partner and “Pass \$2.”

Dependent variables.

As in Studies 1a, 1b, and 2, our main dependent variable was trusting behavior, measured by the decision to pass money in the trust game. Our measures of attitudinal trust and benevolence were identical to the measures we used in Study 2 (r 's $> .86$, $\alpha = .91$). We made a slight revision to our measure of perceived deception to fit the new version of the deception game. Specifically, we asked participants to indicate their agreement with the following statements: “This person sent an honest message about the number chosen by the random number generator as a Sender in The Number Game,” and

“This person lied about the number chosen by the random number generator in The Number Game;” ($r(312) = .86$).

After participants submitted their responses, we asked two recall questions, collected demographic information and asked participants what they thought the purpose of the study was. Participants received bonus payment based upon their decisions in the trust game before they were dismissed.

Results

We report the results from 312 participants (62.8% female; $M_{age} = 21$ years, $SD = 2.50$) who passed all comprehension checks and completed the entire study; 25 participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Passing in the trust game. We first conducted a logistic regression on trusting behavior, using *Deception*, *Intentions*, and the *Deception x Intentions* interaction as independent variables. We found a main effect of *Intentions* ($b = .498, p < .01$), such that participants were more trusting of individuals who made altruistic decisions. Specifically, 47% of participants trusted their partners in the *Altruistic* conditions, whereas only 25% of participants trusted their partners in the *Selfish* conditions, $\chi^2(1, N = 312) = 16.70, p < .01$. We found no main effect of *Deception* and we found no significant *Deception x Intentions* interaction ($ps > .79$). In Figure 3, we display the percentage of participants who passed money in each of the four experimental conditions (*Altruistic Lie*, *Selfish Lie*, *Altruistic Honesty*, and *Selfish Honesty*).

Figure 3 about here

Attitudinal Trust. As in our previous studies, our behavioral and attitudinal measures of trust were highly correlated, $r(312) = .72, p < .001$. A two-way ANOVA revealed a main effect of *Intentions*, $F(1,308) = 78.74, p < .001$, and *Deception*, $F(1,308) = 5.58, p = .02$, on attitudinal trust. Participants trusted their partners more in the *Altruistic* conditions ($M = 4.07, SD = 1.79$) than they did in the *Selfish* conditions ($M = 2.43, SD = 1.49$) and they trusted their partner more in the *Honesty* conditions ($M = 3.46, SD = 1.82$) than in the *Lie* conditions ($M = 3.05, SD = 1.85$). Although lying did not significantly influence behavioral trust, it did influence attitudinal trust. We find no significant interaction between *Deception x Intentions*, $F(1,308) = .19, p = .66$.

Perceived Benevolence. A two-way ANOVA revealed a main effect of *Intentions*, $F(1,308) = 108.70, p < .001$, and *Deception*, $F(1,308) = 18.90, p < .01$, on perceived benevolence. Participants perceived their partner to be more benevolent in the *Altruistic* conditions ($M = 4.82, SD = 1.22$) than in the *Selfish* conditions ($M = 3.42, SD = 1.21$) and to be more benevolent in the *Honesty* conditions ($M = 4.36, SD = 1.27$) than in the *Lie* conditions ($M = 3.89, SD = 1.49$). We find no significant interaction between *Deception x Intentions*, $F(1,308) = .76, p = .36$.

Perceived Deception. Consistent with our manipulation, participants also perceived their partner to be more deceptive in the *Lie* conditions ($M = 6.06, SD = 1.30$) than in the *Honesty* conditions ($M = 2.06, SD = 1.41$), $F(1,255) = 680.02, p < .001$. We find no effect of *Intentions*, $F(1,308) = 1.54, p = .22$, and we find no significant *Deception x Intentions* interaction, $F(1,308) = .28, p = .59$.

Mediation Analyses.

We conducted a moderated mediation analysis using the bootstrap procedure (Hayes, 2013; Preacher, Rucker, & Hayes, 2007) to test the process by which lying and intentions influence trusting behavior.

We predicted that altruistic (and selfish) intentions would influence trusting behavior, regardless of whether the target lied, and that this would be mediated by perceived benevolence. Our mediation model included *Intentions* as the independent variable, *Deception* as the moderator variable, Perceived Benevolence and Perceived Deception as the mediator variables, and Trusting Behavior as the dependent measure. Consistent with our hypothesis, we find that Perceived Benevolence mediates in the expected direction in both the *Lie* conditions (Indirect Effect = 1.14, $SE = .25$; 95% CI [0.70, 1.67]), and the *Honesty* conditions (Indirect Effect = .97, $SE = .23$; 95% CI [0.58, 1.44]), and Perceived Deception does not mediate (both confidence intervals for the indirect effect include zero). These results are unchanged when we use Attitudinal Trust, rather than Trusting Behavior, as the dependent measure. Taken together, these results indicate that perceived benevolence, and not perceived deception, influences trust. That is, deception does not harm trust; selfishness does. We present additional regression analyses in Table 3.

Discussion

In Study 3a, Altruistic individuals were trusted far more than selfish individuals, and this was true whether or not the counterpart's claims were honest or deceptive. Controlling for intentions, we find no direct effect of lying on trusting behavior in either study. This is true even though lying is perceived as deceptive. We use moderated

mediation analysis and confirm that perceived benevolence is the primary mechanism linking prosocial lying with increased trust. Interestingly, trust built on perceived benevolence is not diminished by dishonest acts.

Study 3b

In Study 3b, we extend our investigation by including two control conditions in our experiment. By including control conditions, we can disentangle the beneficial effects of altruistic behavior from the harmful effects of selfish behavior. In our control conditions, participants did not learn about the Sender's decision in the deception game.

Method

Participants. For our 12 cell design, we recruited 1000 participants to participate in an online study in exchange for payment via Amazon Mechanical Turk.

Procedure and Materials. Study 3b was similar to Study 3a, with three notable changes. First, we added two control conditions to disentangle the effects of altruism in increasing trust from the effects of selfishness in decreasing trust. In the control conditions, participants did not learn about the Sender's decision in the deception game.

Second, for simplicity and ease of comprehension we used the Coin Flip game rather than the Number Game for our manipulation of deception. Third, we counterbalanced the order of our behavioral trust measure and our attitudinal trust measure.

In Study 3b, we randomly assigned participants to one of twelve experimental conditions in a 2(Payment Structure: Altruistic Lying-Selfish Honesty vs. Selfish Lying-Altruistic Honesty) x 3(Intentions: Altruistic, Selfish, Control) x 2(Order of measures:

behavior first vs. attitudes first) between-subjects design. Participants learned that the Coin Flip Game had one of two possible payment structures. As in Study 3a, these payment structures enabled us to manipulate whether deception or honesty was associated with selfish or altruistic intentions. The payment structures were identical to those we used in Study 3a. The first payment structure reflected the choice between Altruistic Lying and Selfish Honesty, and the second payment structure reflected the choice between Selfish Lying and Altruistic Honesty.

Therefore, participants learned that the Sender either made the *Altruistic* decision (which was associated with *Lying* or *Honesty*), made the *Selfish* decision (which was associated with *Lying* or *Honesty*), or participants did not learn the Sender's decision (the control conditions). Half of the participants in the control condition learned that the Coin Flip Game reflected the choice between altruistic lying and selfish honesty (the first payment structure) and half learned that the Coin Flip Game reflected the choice between selfish lying and altruistic honesty (the second payment structure). In both control conditions, participants did not learn the Sender's choice.

We refer to these six experimental conditions as *Altruistic Lie*, *Selfish Lie*, *Altruistic Honesty*, *Selfish Honesty*, *Control 1* (learned about the Altruistic Lie-Selfish Honesty payment structure, but did not learn about the Sender's choice), and *Control 2* (learned about the Selfish Lie-Altruistic Honesty payment structure, but did not learn about the Sender's choice).

After participants learned about the Coin Flip Game [and the Sender's decision], participants played a trust game with the Sender.

The Trust Game. We referred to the trust game as “The Choice Game” in the experiment. “The Choice Game” was similar to the trust games we used in our previous studies. Participants had the choice to either “Keep \$1” or trust their partner and “Pass \$1” in the trust game. If participants passed \$1, the amount grew to \$2.50 and their partner had the opportunity to keep \$2.50 or return half (\$1.25).

As in our previous studies, participants had to pass a comprehension check to complete the study.

Dependent variables.

Our primary dependent variable was trusting behavior, measured by the decision to pass money in the trust game. Our measures of attitudinal trust, benevolence, and deception were identical to the measures we used in Study 3a ($r = .93$, α 's $> .88$). However, we did not measure perceived deception in the control conditions because participants did not have any information about whether or not the Sender had deceived their partner.

After participants submitted their responses, we collected demographic information and asked participants what they thought the purpose of the study was. We paid participants a bonus payment based upon their outcome in the trust game before we dismissed them.

Results

We report the results from 974 participants (40.2% female; $M_{age} = 31$ years, $SD = 10.36$) who passed the comprehension checks and completed the entire study; 26 participants failed the comprehension check at the start of the experiment and were automatically eliminated from the study. None of our main results are affected by

question order, and we present our analyses collapsed across this factor. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Passing in the trust game. We first conducted a logistic regression on trusting behavior, using *Payment Structure*, *Intentions*, and the *Payment Structure x Intentions* interaction as independent variables. We found a main effect of *Intentions*, ($b = .938$, $p < .001$); participants were significantly more likely to pass money in the trust game in the *Altruistic* conditions (69%) than in the *Control* conditions (47%); $\chi^2(1, N = 654) = 32.10$, $p < .01$, and in the *Selfish* conditions (25%), $\chi^2(1, N = 650) = 121.43$, $p < .01$. Participants were also significantly more likely to trust their partner in the *Control* conditions than they were in the *Selfish* conditions, $\chi^2(1, N = 644) = 32.53$, $p < .01$.

We found no effects of *Payment Structure*, nor did we find a significant *Intentions x Payment Structure* interaction ($ps > .86$). In Figure 4, we display the percentage of participants who passed money in each of the six experimental conditions.

Figure 4 about here

Attitudinal Trust. As in our previous studies, behavioral and attitudinal measures of trust were highly correlated, $r(974) = .71$, $p < .001$, and followed a similar pattern. A two-way ANOVA revealed a significant main effect of *Intentions*, $F(2,968) = 240.74$, $p < .001$, such that participants trusted their partners more in the *Altruistic* conditions ($M = 4.70$, $SD = 1.61$) than the *Control* conditions ($M = 3.22$, $SD = 1.78$), $t(653) = 11.86$, $p < .001$; and the *Selfish* conditions ($M = 1.96$, $SD = 1.37$), $t(649) = 21.94$, $p < .001$.

Participants were also more trusting of their partner in the *Control* conditions than in the *Selfish* conditions, $t(643) = 10.00, p < .001$.

We found no main effect of *Payment Structure*, $F(1, 968) = 0.25, p = .62$. There was, however, a significant *Intentions x Payment Structure* interaction, $F(2, 968) = 4.30, p < .05$. Participants trusted individuals who told selfish lies ($M = 1.73, SD = 1.09$) significantly less than individuals who were selfishly honest ($M = 2.18, SD = 1.56$), $t(319) = 2.54, p = .01$, but we found no difference in trust between individuals who told altruistic lies ($M = 4.68, SD = 1.64$) and individuals who were altruistically honest ($M = 4.71, SD = 1.58$), $t(329) = 0.17, p = .87$. We also found no difference in trust between the two control conditions ($M = 3.08, SD = 1.66$ vs. $M = 3.35, SD = 1.89$), $t(323) = 1.53, p = .13$. These results suggest that deception in the service of altruism does not undermine trust, but that deception in the service of selfishness does harm trust.

Perceived Benevolence. Perceived benevolence followed the same pattern as attitudinal trust. A two-way ANOVA revealed a significant main effect of *Intentions*, $F(2,968) = 377.80, p < .001$, such that participants perceived their partner to be more benevolent in the *Altruistic* conditions ($M = 5.20, SD = 1.01$) than they did in the *Control* conditions ($M = 4.29, SD = 0.98$), $t(653) = 11.18, p < .001$, and the *Selfish* conditions ($M = 2.98, SD = 1.20$), $t(649) = 27.24, p < .001$. Participants also rated their partners as more benevolent in the *Control* conditions than they did in the *Selfish* conditions, $t(643) = 16.20, p < .001$.

We also found a main effect of *Payment Structure*, $F(1, 968) = 20.01, p < .001$; partners who faced the opportunity to tell altruistic lies were perceived to be more benevolent ($M = 4.30, SD = 1.32$) than were partners who faced the opportunity to tell

selfish lies ($M = 4.04, SD = 1.47$). This effect was qualified by a significant *Intentions x Payment Structure* interaction, $F(2, 968) = 17.03, p < .001$. Participants rated partners who told selfish lies ($M = 2.54, SD = 1.02$) to be significantly less benevolent than partners who were selfishly honest ($M = 3.39, SD = 1.22$), $t(319) = 7.28, p < .001$, but we found no difference in perceived benevolence between partners who told altruistic lies ($M = 5.25, SD = 1.07$) and partners who were altruistically honest ($M = 5.15, SD = 0.94$), $t(329) = 0.91, p = .36$. In other words, selfish deception was perceived to be particularly malevolent. There was no difference in perceived benevolence between the two control conditions ($M = 4.27, SD = 0.92$ vs. $M = 4.32, SD = 1.04$), $t(323) = 0.46, p = .65$.

Perceived Deception. Consistent with our manipulation, a two-way ANOVA revealed a significant *Intentions x Payment Structure* interaction, $F(1, 645) = 1611.15, p < .001$, such that altruistic lies were perceived to be more deceptive ($M = 5.17, SD = 1.33$) than selfish honesty ($M = 2.77, SD = 1.53$), $t(324) = 18.46, p < .001$, and selfish lying was perceived to be more deceptive ($M = 6.47, SD = 0.88$) than altruistic honesty ($M = 1.51, SD = 0.76$), $t(323) = 38.15, p = .001$.

We also found a main effect of *Intentions*, $F(1, 645) = 195.15, p < .001$, such that selfishness was perceived to be more deceptive ($M = 4.57, SD = 1.07$) than altruism ($M = 3.29, SD = 2.13$). In other words, the same lie was perceived to be more deceptive when it was associated with selfish, rather than altruistic, intentions. We found no main effect of *Payment Structure*, $F(1, 645) = 0.08, p = .78$.

Discussion

In Study 3a, we demonstrate that deception itself has no effect on benevolence-based trust. In Study 3b, we include control conditions and document both a penalty for

selfishness and a benefit for altruism. Selfish intentions, whether they were associated with honesty or deception, harmed trust; altruistic intentions, whether they were associated with honesty or deception, increased trust.

Although we find no differences between altruistic lies and altruistic honesty, in Study 3b, we do find that selfish lies are penalized relative to selfish honesty. Individuals may perceive honesty as the default decision, whereas lying may reflect a willful departure that is more diagnostic of intentionality. In this case, lying to reap selfish benefits may convey a stronger signal of malevolent intentions than honesty that yields the same outcome.

Study 4

In our previous studies, we demonstrate that prosocial lies can increase trust. We measure trust using the trust game, and conceptualized trust as the willingness to be vulnerable to another person when there is an opportunity for exploitation. In Study 3a we demonstrate that trust behavior and attitudes are mediated by perceptions of benevolence and are largely unaffected by deception. In other words, prosocial deception increases benevolence-based trust.

Benevolence-based trust is important in close relationships (e.g., Kim et al., 2006) and likely influences behaviors such as the decision to loan money or property to another person, the decision to rely on someone for emotional support, or the decision to share sensitive information with someone (e.g., McAllister, 1995; Currall & Judge, 1995; Glaeser et al., 2000; Levin & Cross, 2004). However, there are many other types of trust that pervade our social interactions. The people we trust to be kind may not be the same

people we trust to avoid temptation and act with integrity. In Study 4, we demonstrate how prosocial lies influence different types of trust. Specifically, we explore how prosocial lies influence integrity-based trust.

Integrity-based trust reflects the belief that the trustee adheres to a set of acceptable ethical principles, such as honesty and truthfulness (Butler & Cantrell, 1984; Mayer et al., 1995; Kim et al., 2004). Integrity may be particularly important when trust decisions hinge on issues of veracity. For example, in some cases individuals need to rely on information or advice. In these cases, integrity may matter more than benevolence and prosocial lies will decrease trust. In Study 4 we demonstrate that prosocial lies increase benevolence-based trust, but harm integrity-based trust.

Pilot Study

We introduce a new trust game, the Rely-or-Verify game, to capture integrity-based trust. We report results from a pilot study to demonstrate that decisions within the game corresponded with integrity-based trust.

The Rely-or-Verify Game. We designed the Rely-or-Verify game to parallel the decision to trust a counterpart's claim. For example, an employer faces the decision of whether or not to trust a prospective employee's claim about their work experience. The employer could either trust the prospective employee's claim or verify the claim, at a cost. Similarly, negotiators, relational partners, and parents can either trust or verify the claims their counterparts make.

The following features characterize the Rely-or-Verify game: First, the trustee derives a benefit from successful deception (e.g., by over-stating prior work experience). Second, the truster cannot distinguish deception from honesty without verifying a claim.

Third, for the truster, relying on the trustee's claim is risky, and fourth, verifying a claim is costly. The decision to rely on a potential trustee's information reflects the willingness to be vulnerable predicated upon honesty. This decision operationalizes Rotter's (1971: p. 444) original definition of trust: "a generalized expectancy...that the word, promise, verbal, or written statement of another individual or group can be relied on." In contrast to passing money in the trust game, which depends on perceived benevolence, the decision to rely on a counterpart's claim in Rely-or-Verify depends on perceptions of the counterpart's reputation for honesty and truth (Butler & Cantrell, 1984). That is, this decision reflects integrity-based trust.

In Rely-or-Verify, Player 1 (the trustee) makes a claim that is either accurate or inaccurate. Player 2 observes the claim and decides to either **Rely** (trust) or **Verify** (not trust) the claim. If Player 1's claim is **inaccurate** and Player 2 **relies** on the claim, Player 1 earns a_1 and Player 2 earns a_2 . If Player 1's claim is **inaccurate** and Player 2 **verifies** it, Player 1 earns b_1 and Player 2 earns b_2 . If Player 1's claim is **accurate** and Player 2 **relies** on it, Player 1 earns c_1 and Player 2 earns c_2 . If Player 1's claim is **accurate** and Player 2 **verifies** it, Player 1 earns d_1 and Player 2 earns d_2 .

The payoffs for Player 1 are structured such that $a_1 > c_1 \geq d_1 > b_1$. For Player 1, deception is risky; for Player 1, deception yields the highest payoff if Player 2 relies on the deceptive claim, but it yields the lowest payoff if Player 2 verifies the deceptive claim.

The payoffs for Player 2 are structured such that $c_2 > d_2 \geq b_2 > a_2$. In other words, Player 2 earns the highest payoff by relying on accurate information and the lowest payoff by relying on inaccurate information. Verification is costly, but minimizes risk.

By verifying information, Player 2 learns the truth. Thus, verification yields the same outcome for Player 2, regardless of whether or not Player 1 told the truth. We depict the general form of Rely-or-Verify in Figure 5.

Figure 5 about here

In our experiment, we term Player 1 the “Red Player” and Player 2 the “Blue Player.” The Red Player sends a message to the Blue Player. In this case, the Red Player had to report whether or not the amount of money in a jar of coins was odd or even. The Blue Player received this message and could either Rely on the message or Verify the message. We provide the full instructions and the exact game we used in Appendix A.

Participants. We recruited 198 participants from a city in the northeastern United States to participate in a pilot study of Rely-or-Verify in exchange for a \$10 show-up fee.

Method. Participants in the pilot study read the full instructions of the Rely-or-Verify game (see Appendix A) and were assigned to the role of the “Blue Player.” Participants had to pass a comprehension check in order to complete the entire study. Participants who failed the comprehension check twice were automatically removed from the experiment.

Participants who passed the comprehension check received a message from a confederate “Red Player,” informing them that the amount of money in the jar was either odd or even. The decision to “Rely” represents our behavioral measure of integrity-based trust.

After participants made a decision to Rely or Verify, they rated how much they trusted their partner, and rated their partner's benevolence and integrity. We measured trusting attitudes using three items ($\alpha = .84$): "I trust my partner," "I am willing to make myself vulnerable to my partner," and "I am confident that my partner sent me an accurate message;" 1 = "Strongly disagree" and 7 = "Strongly agree." We measured perceived benevolence using the same scale we used in Studies 3a and 3b ($\alpha = .78$), and we measured perceived integrity using three items ($\alpha = .66$): "This person has a great deal of integrity," "I can trust this person's word," and "This person cares about honesty and truth;" 1 = "Strongly disagree" and 7 = "Strongly agree."

After participants made Rely-or-Verify decisions and rated their partner, they answered demographic questions, were paid, and dismissed.

Results. Nearly all of the participants (98%) passed the comprehension check and completed the entire study. A total of 31.3% of participants chose "Rely" and trusted their partner. We did not identify any gender differences in behavior.

Importantly, the decision to "Rely" was closely related to perceptions of trustworthiness, $r(194) = .71, p < .001$. Trusting behavior in Rely-or-Verify was correlated with both perceived benevolence, $r(194) = .48, p < .001$, and perceived integrity $r(194) = .52, p < .001$. That is, the decision to trust a target's claim, rather than verify it, reflects the belief that the target would not purposefully hurt others (benevolence) *and* the belief that the target cares about honesty and truth (integrity).

Main Study

In our main study, participants learned about a counterpart who had either told prosocial lies or who had been honest in a series of prior interactions. After learning this

information, participants played either the Trust Game or the Rely-or-Verify game with the counterpart.

Method

Participants. We recruited 500 participants to participate in an online study in exchange for payment via Amazon Mechanical Turk.

Procedure and Materials. Participants in Study 4 learned about a series of decisions a confederate counterpart made as a Sender in the Coin Flip Game. This was the same Coin Flip Game we used in Studies 1a, 1b, 2, and 3b. Participants then played either the Trust Game or the Rely-or-Verify game with this counterpart. We randomly assigned participants to one of four cells from a 2(Deception: Prosocial lie vs. Honesty) x 2(Game: Trust Game vs. Rely-or-Verify) between-subjects design.

In Study 4, participants learned that the Sender had played the Coin Flip Game four times with four different partners. We altered the payoffs associated with deception in each of the four rounds of the game so that we could include both altruistic and mutually beneficial lies in a single manipulation. By using repeated behavior to manipulate prosocial deception, we strengthened our manipulation. This manipulation made it clear that the Sender was either committed to honesty (telling the truth even when it was costly) or to benevolence (helping the Receiver even when it was costly).

Participants learned about four decisions the Sender had made in four rounds of The Coin Flip Game. In rounds 1 and 3, the Sender faced the choice between an altruistic lie and selfish honesty. In rounds 2 and 4, the Sender faced the choice between a mutually beneficial lie and mutually harmful honesty. Participants learned that the Sender made one of the following two profiles of decisions: Prosocial Lies {Altruistic lie, mutually

beneficial lie, altruistic lie, mutually beneficial lie} or Honesty {Selfish truth, mutually harmful truth, selfish truth, mutually harmful truth}. We include the payoffs associated with each choice in Table 4.

After participants learned about the Sender's four decisions, participants played either the trust game or the Rely-or-Verify game with the Sender. The Trust Game we used was identical to the version of the trust game we used in Study 3b. The version of the Rely-or-Verify game we used was identical to the version we used in the pilot study.

Dependent variables.

Our main dependent variable was trusting behavior, measured by the decision to pass money in the trust game or "Rely" in the Rely-or-Verify game. Our measures of attitudinal trust for Rely-or-Verify were identical to the measures we used in the pilot study. We adapted the wording of these items to create a parallel measure of attitudinal trust for the trust game ($\alpha = .92$). We provide all items and anchors in Appendix B.

We measured perceived deception, using the same measures we used in our previous studies ($\alpha = .94$). We measured perceived benevolence as we did before, but to be sure to distinguish benevolence from integrity, we eliminated the item, "This person has good intentions;" $r(457) = .72, p < .001$.

After participants submitted their responses, we asked a recall question, collected demographic information and asked participants what they thought the purpose of the study was. The next day, we followed up with participants to pay them a bonus payment based on their decisions in our study.

Results

We report results from 457 participants (31.6% female; $M_{\text{age}} = 31$ years, $SD = 9.87$) who passed all comprehension checks and completed the entire study; 43 participants failed the comprehension check and were automatically removed from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 5.

Trusting behavior. We first conducted a logistic regression on trusting behavior using *Deception*, *Game*, and the *Deception* x *Game* interaction as independent variables. Importantly, we found a significant *Lying* x *Game* interaction; $b = .37, p < .01$. In the trust game, participants were more likely to pass money to their partners in the *Prosocial Lie* condition (57%) than in the *Honesty* condition (40%), $\chi^2(1, N = 262) = 7.41, p < .01$. We find the opposite pattern of results in the Rely-or-Verify game; participants playing Rely-or-Verify were *less* likely to rely on their partners in the *Prosocial Lie* condition (37%) than in the *Honesty* condition (57%); $\chi^2(1, N = 195) = 7.75, p < .01$. In other words, prosocial lying increased benevolence-based trust and harmed integrity-based trust. We found no main effect of *Deception* or *Game* ($ps > .73$). We depict these results in Figure 6.

Figure 6 about here

Attitudinal Trust. Our attitudinal trust measures parallel our behavioral trust results. We find that trusting attitudes were highly correlated with trusting behavior in both games, each $r \geq .80$ (see Table 5).

A two-way ANOVA revealed a significant *Deception* x *Game* interaction, $F(1,453) = 17.57, p < .001$, such that prosocial lying increased trusting attitudes in the Trust Game, but decreased trusting attitudes in the Rely-or-Verify game.

Specifically, participants trusted the prosocial liar more than the honest individual in the *Trust Game* conditions ($M = 4.11, SD = 2.08$ vs. $M = 3.54, SD = 1.86$), $t(261) = 2.48, p = .014$, but trusted the prosocial liar less than the honest individual in the *Rely-or-Verify* conditions ($M = 3.57, SD = 1.79$ vs. $M = 4.46, SD = 1.56$), $t(194) = 3.38, p < .01$. We did not find a significant main effect of *Deception*, $F(1,453) = 1.21, p = .27$, or *Game*, $F(1,453) = .89, p = .34$.

Perceived Benevolence. Ratings of perceived benevolence followed a similar pattern. A two-way ANOVA revealed a significant *Deception* x *Game* interaction, $F(1,453) = 5.93, p = .015$, but no main effect of *Deception*, $F(1,453) = 1.89, p = .17$, or *Game*, $F(1,453) = .15, p = .70$. Specifically, participants judged the prosocial liar to be more benevolent than the honest individual in the *Trust Game* conditions ($M = 4.72, SD = 1.74$ vs. $M = 4.16, SD = 1.53$), $t(261) = 2.92, p < .01$, but there was no difference between the prosocial liar and the honest individual in the *Rely-or-Verify* conditions ($M = 4.30, SD = 1.51$ vs. $M = 4.46, SD = 1.32$), $t(194) = 0.70, p = .48$. It is possible that individuals did not rate the prosocial liar as more benevolent in the Rely-or-Verify game because their ratings were influenced by their prior trust decisions.

Perceived Deception. As expected, individuals who told prosocial lies were perceived to be more deceptive ($M = 5.81, SD = 1.17$) than individuals who were honest ($M = 1.75, SD = 1.11$), $F(1, 453) = 1393.2, p < .001$. We did not find a main effect of

Game, $F(1,453) = .60$, $p = .44$, or a significant *Deception* x *Game* interaction, $F(1,453) = .04$, $p = .84$.

Discussion

In our prior studies, we documented the beneficial effects of prosocial lies on trust. In Study 4, we use a repeated deception game and demonstrate that prosocial lies increase benevolence-based trust, but harm integrity-based trust. That is, we reveal that prosocial lies are not universally beneficial; prosocial lies affect different types of trust differently.

To measure different types of trust, we introduce a new tool. The prevailing behavioral measure of trust, the trust game, reflects benevolence-based trust. We introduce the Rely-or-Verify game to measure integrity-based trust.

General Discussion

Across four studies, we demonstrate that lying can increase trust. In particular, we find that prosocial lies, false statements told with the intention of benefitting others, increase benevolence-based trust. In Study 1a, participants trusted counterparts more when the counterpart told them an altruistic lie than when the counterpart told the truth. In Study 1b, we replicate our result and rule out direct reciprocity as an alternative mechanism. In this study, participants observed, rather than experienced deception.

In Studies 2, 3a, and 3b, we examine different types of lies. We find that participants trusted individuals who told non-altruistic, prosocial lies and mutually beneficial lies more than individuals who told truths that harmed others. Our findings reveal that benevolence, demonstrating concern for others, is far more important for fostering some types of trust, than either honesty or selflessness. In fact, we find that

deception per se, does surprisingly little to undermine trust in the trust game. We disentangle the beneficial effects of benevolence from the harmful effects of selfishness and find that compared to a control condition, benevolence boosts trust and selfishness harms trust.

In Study 4, we investigate how prosocial lying influences distinct types of trust. The trust game reflects benevolence-based trust, and we introduce a new game, the Rely-or-Verify game to capture integrity-based trust. We demonstrate that prosocial lying increases benevolence-based trust, but undermines integrity-based trust. That is, prosocial lies do not always benefit trust; the same actions can have divergent effects on benevolence-based and integrity-based trust.

Contributions and Implications

Our findings demonstrate that the relationship between deception and trust is far more complicated than prior work has assumed. Lying, per se, does not always harm trust.

In prior work, deception has been unfairly disparaged because scholars have conflated deception with self-serving intentions. We find that although deception can exacerbate the negative inferences associated with selfish actions, deception does not undermine the positive inferences associated with prosocial actions.

Our research contributes to the deception and trust literatures in three ways. First, we highlight the importance of studying a broader range of deceptive behaviors. Prosocial lying is pervasive, but we know surprisingly little about the interpersonal consequences of prosocial lies. Although most research assumes that deception is

harmful, we document potential benefits of deception. Prosocial lies signal benevolence and can increase trust.

Second, we provide insight into the antecedents of trust. Trust scholars have assumed that both integrity and benevolence are antecedents of trust, yet little research investigates when each of these values matters. Our research suggests that benevolence may be the primary concern for many trust decisions. We are the first to independently manipulate benevolence and honesty in this context and draw causal inferences about how they each impact trust.

Third, we demonstrate that identical actions can have divergent effects on different trust decisions. Scholars have used the term “trust” to refer to a broad range of behaviors. For example, trust has been used to describe the willingness to hire someone (Kim et al., 2004), to give someone responsibility without oversight (Kim et al., 2004; Mayer & Davis, 1999), to rely on someone’s word (Johnson-George & Swap, 1982; Rotter, 1971), and to expose oneself to financial risk (Berg et al., 1995; Pillutla et al., 2003; Schweitzer et al., 2006; McCabe et al., 2003; Glaeser et al., 2000; Malhotra & Murnighan, 2002; Malhotra, 2004). Our findings suggest that different types of trust may guide these decisions, and the same information may influence these decisions in different ways.

Our research has both methodological and managerial implications. We introduce a new tool that researchers can use to capture the central elements of trusting acts. The Rely-or-Verify game is similar to the trust game, but it offers distinct advantages. First, the Rely-or-Verify game reflects perceptions of both benevolence and integrity, rather than benevolence alone. Second, in contrast with the trust game, the truster in the Rely-

or-Verify game moves second. That is, in contrast to Player 1's decision in the trust game (who might pass to engender reciprocity; Chou, Halevy, & Murnighan, 2011), there is no strategic signal in Player 2's decision to trust in Rely-or-Verify. In addition, because trust in the Rely-or-Verify game is not characterized by the decision to share money, trust decisions in the Rely-or-Verify game are also unlikely to reflect social preferences for fairness or altruism, a concern that scholars have raised about the trust game (e.g., Ashraf et al., 2006).

Prescriptively, our findings suggest that we should reconsider how we characterize deception. Parents, leaders and politicians often publicly and emphatically denounce lying—even though they often engage in it (Nyberg, 1993; Heyman, Luu, Lee, 2009; Grover, 2005). Acknowledging the benefits of prosocial lies could free individuals of this hypocrisy. In fact, authority figures could explicitly embrace certain types of deception and teach others when and how to lie. This would reflect a stark contrast to the current practice of asserting that lying is universally wrong, while modeling that it is often right.

Managers should also consider if honesty is always the best policy. Honesty, although often considered a virtue, in some cases may be selfish and mean-spirited. In many conversations, individuals make a trade-off between being honest and being kind. In order for employees to engender trust, benevolence may be far more important than honesty.

Limitations and Future Directions

In our studies, we experimentally manipulated behavior in the deception game, which afforded us experimental control. By altering the monetary payoffs associated with

honesty and lies, we were able to send unambiguous signals about the intentions associated with each lie. This enables us to draw causal inferences about how prosocial intentions and deception differentially influence distinct forms of trust.

However, many prosocial lies are characterized by features that we did not capture in our experiments. For example, we study lies that generated monetary gains. Although some lies generate monetary outcomes, many lies, and prosocial lies in particular, are motivated by the desire to protect people's feelings (DePaulo, 1992). These lies may be perceived to be more innocuous and be more likely to foster emotional security, an important component of trust in close relationships (Rempel et al., 1985). Furthermore, lies told to avoid losses may be perceived to be more benevolent than lies told to accrue gains. Avoiding a loss is often much more psychologically powerful than generating a gain (Kahneman & Tversky, 1979), and thus, deceived parties may be particularly grateful to be the beneficiaries of these types of lies.

In our study, the motives and outcomes associated with deception were clear. However, in practice, both motives and the link between acts and outcomes may be difficult to gauge. In some cases, people attribute selfish motives to prosocial acts (Critcher & Dunning, 2011; Fein, 1996; Newman & Cain, 2014; Miller, 1999; Lin-Healy & Small, 2013). This may be particularly true for prosocial acts that violate other moral principles, such as the principle of honesty. Although participants in our studies were not distrusting of individuals who told lies to help themselves *and* others, it is possible that if the selfish gains associated with mutually beneficial lies were sufficiently large, observers would discount the liar's prosociality.

There are likely contexts, however, in which lies told with good intentions are not perceived to be benevolent. For example, Wang and Murnighan (2013) found that some lies that help others – such as lying to a patient about a medication – are perceived to be less benevolent than honesty, and consequently, harm trust. We expect that the extent to which a lie is perceived as benevolent is immensely important. Future research should explore when and why favorable attributions are made for ethically ambivalent actions.

Quite possibly, many factors that we did not explore will moderate the relationship between prosocial deception and trust. For example, the stage of a relationship may matter. In early stages of relationship development, emotional security may be a primary concern, and prosocial lying may be particularly beneficial. In late stages of relationships, honesty may be a stronger signal of intimacy than kindness. Perhaps as relationships develop, the role of prosocial lying will change. It is also possible that prosocial lies have detrimental long-term consequences. If an individual develops a reputation for dishonesty, prosocial lies may become less credible. Future work should explore trust as a dynamic construct (Lewicki et al., 2006; Schoorman, Mayer, & Davis, 2007; Schweitzer et al., 2006).

It is possible that our attitudes towards deception do not reflect intrinsic preferences for honesty and truth, but instead reflect our expectations of different relational partners. We may expect people in some roles to support and help us, but expect others to be objective and provide us with accurate information. Understanding how the nature of trust differs across relationships is an important next step for trust research.

Gender and power may also influence our preferences for honesty and kindness. For example, women tell more prosocial lies than men (Erat & Gneezy, 2012) and are generally expected to be more polite than men (Brown & Levinson, 1987). Although we identified no gender differences in our studies, there may be circumstances in which women suffer greater backlash for impolite honesty than men. This may also be the case for low-power individuals who are expected to conform to politeness norms (Brown & Levinson, 1987). Sanctions for impolite honesty may have detrimental consequences in organizations by curbing the flow of information and curtailing employee voice.

A final open question is whether prosocial liars are strategic or truly prosocial. Individuals have a strong desire to be viewed favorably by others (Baumeister, 1982; Jones & Wortman, 1973; Leary & Kowalski, 1990; Tetlock, 2002) and telling someone nice things, regardless of one's true beliefs, may be a strategic way to garner future favors. This may be particularly likely when the lie goes undetected.

Conclusion

We challenge the assumption that deception harms trust. Prior studies of deception have confounded lying with selfish intentions. By disentangling the effects of intentions from deception, we demonstrate that deception itself does not always harm trust. Although prosocial lies harm integrity-based trust, they *increase* benevolence-based trust. In many cases, we find that intentions matter far more than veracity.

References

- Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, 9(3), 193-208.
- Atwater, L. E. (1988). The relative importance of situational and individual variables in predicting leader behavior: The surprising impact of subordinate trust. *Group & Organization Management*, 13(3), 290-310.
- Baccara, M., & Bar-Isaac, H. (2008). How to organize crime. *The Review of Economic Studies*, 75(4), 1039-1067.
- Barney, J. B., & Hansen, M. H. (1994). Trustworthiness as a source of competitive advantage. *Strategic Management Journal*, 15(S1), 175-190.
- Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin*, 91, 3-26.
- Bazerman, M. H. 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(1), 122-142.
- Bhattacharya, R., Devinney, T. M., & Pillutla, M. M. (1998). A formal model of trust based on outcomes. *Academy of Management Review*, 23(3), 459-472.
- Bies, R. J., & Tripp, T. M. (1996). Beyond distrust: "Getting even" and the need for revenge. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in Organizations: Frontiers of theory and research*: 246-260. Thousand Oaks, CA: Sage.
- Blau, P. M. (1964). *Exchange and power in social life*. Piscataway, NJ: Transaction Publishers.
- Bok, S. (1978). *Lying: Moral choices in public and private life*. New York, NY:

- Pantheon.
- Boles, T. L., Croson, R. T., & Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes*, 83(2), 235-259.
- Bowles, S., & Gintis, H. (2004). Persistent parochialism: trust and exclusion in ethnic networks. *Journal of Economic Behavior & Organization*, 55(1), 1-23.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. New York, NY: Cambridge University Press.
- Broomfield, K. A., Robinson, E. J., & Robinson, W. P. (2002). Children's understanding about white lies. *British Journal of Developmental Psychology*, 20(1), 47-65.
- Butler Jr, J. K., & Cantrell, R. S. (1984). A behavioral decision theory approach to modeling dyadic trust in superiors and subordinates. *Psychological Reports*, 55(1), 19-28.
- Butler, J. K. (1991). Toward understanding and measuring conditions of trust: Evolution of conditions of trust inventory. *Journal of Management*, 17(3), 643-663
- Carr, A. (1968). Is business bluffing ethical?. *Harvard Business Review*, 46, 143-155.
- Chou, E. Y., Halevy, N., & Murnighan, J. K. (2011, June). Trust as a tactic: The calculative induction of reciprocity. In *IACM 24th Annual Conference Paper*.
- Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology*, 45(6), 1321-1324.

- Critcher, C. R., & Dunning, D. (2011). No good deed goes unquestioned: Cynical reconstruals maintain belief in the power of self-interest. *Journal of Experimental Social Psychology, 47*, 1207–1213.
- Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior & Organization, 51*(2), 143-159.
- Currall, S. C., & Judge, T. A. (1995). Measuring trust between organizational boundary role persons. *Organizational behavior and Human Decision processes, 64*(2), 151-170.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111*(2), 203.
- DePaulo, B. M., & Bell, K. L. (1996). Truth and investment: Lies are told to those who care. *Journal of Personality and Social Psychology, 71*(4), 703-716.
- DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74*(1), 63-79.
- Dirks, K. T. (2000). Trust in leadership and team performance: Evidence from NCAA basketball. *Journal of applied psychology, 85*(6), 1004-1012.
- Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science, 12*(4), 450-467.
- Dunn, J., Ruedy, N. E., & Schweitzer, M. E. (2012). It hurts both ways: How social comparisons harm affective and cognitive trust. *Organizational Behavior and Human Decision Processes, 117*(1), 2-14.
- Erat, S., & Gneezy, U. (2012). White lies. *Management Science, 58*(4), 723-733.

- Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, *70*, 1164–1184.
- Ferrin, D. L., Kim, P. H., Cooper, C. D., & Dirks, K. T. (2007). Silence speaks volumes: The effectiveness of reticence in comparison to apology and denial for responding to integrity-and competence-based trust violations. *Journal of Applied Psychology*, *92*(4), 893-908.
- Ford, C. V, King, B.H., & Hollender, M.H. (1988). Lies and liars: Psychiatric aspects of prevarication. *American Journal of Psychiatry*, *145*(5), 554-562.
- Giffin, K. 1967. The contribution of studies of source credibility to a theory of interpersonal trust in the communication department. *Psychological Bulletin*, *68*(2), 104-120.
- Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review*, *34*(1), 127-145.
- Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior and Organization*, in press.
- Gino, F., & Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science*, *20*(9), 1153-1160.
- Gino, F., & Pierce, L. (2010a). Robin Hood under the hood: Wealth-based discrimination in illicit customer help. *Organization Science*, *21*(6), 1176-1194.
- Gino, F., & Pierce, L. (2010b). Robin Hood under the hood: Wealth-based discrimination in illicit customer help. *Organization Science*, *21*(6), 1176-1194.

- Gino, F., & Shea, C. (2012). Deception in negotiations: The role of emotions. *Handbook of conflict resolution*. Oxford University Press, New York.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, 115(3), 811-846.
- Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, 95(1), 384-394.
- Goffman, E. (1967). On face-work. *Interaction ritual*, 5-45.
- Golembiewski, R. T., & McConkie, M. (1975). The centrality of interpersonal trust in group processes. In C. L. Cooper (Ed.), *Theories of group processes* (pp. 131-185). London, UK: Wiley.
- Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), 481-510.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), 101-124.
- Grover, S. L. (2005). The truth, the whole truth, and nothing but the truth: The causes and management of workplace lying. *The Academy of Management Executive*, 19(2), 148-157.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Heyman, G. D., Luu, D. H., & Lee, K. (2009). Parenting by lying. *Journal of Moral Education*, 38(3), 353-369.
- Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philosophical ethics. *Academy of Management Review*, 20(2), 379-403.

- Houser, D., Schunk, D., & Winter, J. (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior & Organization*, 74(1), 72-81.
- Iezzoni, L. I., Rao, S. R., DesRoches, C. M., Vogeli, C., & Campbell, E. G. (2012). Survey shows that at least some physicians are not always open or honest with patients. *Health Affairs*, 31(2), 383-391.
- Jones, E. E., & Wortman, C. (1973). *Ingratiation: An attributional approach*. Morristown, NJ: General Learning Press.
- Johnson-George, C., & Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology*, 43(6), 1306-1317.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.
- Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *The American Economic Review*, 95(5), 1688-1699.
- Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence-vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, 99(1), 49-65.
- Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (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(1), 104.

- Koning, L., Steinel, W., Beest, I. V., & van Dijk, E. (2011). Power and deception in ultimatum bargaining. *Organizational Behavior and Human Decision Processes*, 115(1), 35-42.
- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34-47.
- Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, 50(11), 1477-1490.
- Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology*, doi: 10.1016/j.jesp.2014.03.005
- Lewicki, R. J., & Bunker, B. B. (1995). *Trust in relationships: A model of development and decline*. Jossey-Bass.
- Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, 32(6), 991-1022.
- Lewis, M. & Saarni, C. (1993). *Lying and deception in everyday life*. New York, NY: The Guilford Press.
- Lin-Healy, F., & Small, D. A. (2013). Nice guys finish last and guys in last are nice: The clash between doing well and doing good. *Social Psychological and Personality Science*, 4 (6), 693-699.
- Lount Jr, R. & Pettit, N. (2012). The social context of trust: The role of status. *Organizational Behavior and Human Decision Processes*, 117(1), 15-23.

- Lount, R. B., Zhong, C. B., Sivanathan, N., & Murnighan, J. K. (2008). Getting off on the wrong foot: The timing of a breach and the restoration of trust. *Personality and Social Psychology Bulletin*, *34*(12), 1601-1612.
- Malhotra, D. (2004). Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes*, *94*(2), 61-73.
- Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly*, 534-559.
- Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, *84*(1), 123.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734.
- McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24-59.
- McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, *52*(2), 267-275.
- Meier, S., Pierce, L., & Vaccaro, A. (2013). Trust and Parochialism in a Culture of Crime. Washington University of St. Louis: Working Paper.
- Miller, D. T. (1999). The norm of self-interest. *American Psychologist*, *54*(12), 1053-1060.
- Murnighan, J. K. (1991). *The dynamics of bargaining games*. Englewood Cliffs, NJ:

- Prentice Hall.
- Newman, G., & Cain, D. M. (2014). Tainted Altruism: When doing some good is worse than doing no good at all. *Psychological Science*.
- Nyberg, D. (1993). *The varnished truth: Truth telling and deceiving in ordinary life*. Chicago: University of Chicago Press.
- O'Connor, K. M., & Carnevale, P. J. (1997). A nasty but effective negotiation strategy: Misrepresentation of a common-value issue. *Personality and Social Psychology Bulletin*, 23(5), 504-515.
- Pillutla, M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, 39, 448-455.
- Planalp, S., Rutherford, D. K., & Honeycutt, J. M. (1988). Events that increase uncertainty in personal relationships II: Replication and extension. *Human Communication Research*, 14(4), 516-547.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42(1), 185-227.
- Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, 49(1), 95-112.
- Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, 41(4), 574-599.
- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American psychologist*, 26(5), 443-452.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after

- all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393-404.
- Santoro, M. A., & Paine, L. S. (1993). Sears auto centers (Harvard Business School case 9-394-010). *Boston: Harvard Business School Publishing*.
- Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, 32(2), 344-354.
- Schweitzer, M. E., & Croson, R. (1999). Curtailing deception: The impact of direct questions on lies and omissions. *International Journal of Conflict Management*, 10(3), 225-248.
- Schweitzer, M. E. & Gibson, D. E. (2008). Fairness, feelings, and ethical decision-making: Consequences of violating community standards of fairness. *Journal of Business Ethics*, 77(3), 287-301.
- 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.
- Sheppard, B. H., & Sherman, D. M. (1998). The grammars of trust: A model and general implications. *Academy of Management Review*, 23(3), 422-437.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.
- Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic “remedies” for trust/distrust. *Organization Science*, 4(3), 367-392.

- Steinel, W., & De Dreu, C. K. (2004). Social motives and strategic misrepresentation in social decision making. *Journal of Personality and Social Psychology*, 86(3), 419.
- Talwar, V., Murphy, S. M., & Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development*, 31(1), 1-11.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, 109, 451-471.
- Tyler, J. M. & Feldman, R. S. (2006). Truth, lies, and self-presentation: How gender and anticipated future interaction relate to deceptive behavior. *Journal of Applied Social Psychology*, 34(12), 2602-2615.
- Tyler, J. M., Feldman, R. S., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology*, 42(1), 69-77.
- Tyler, J. M. & Feldman, R. S. (2004). Cognitive demand and self-presentation efforts: The influence of situational importance and interaction goal. *Self and Identity*, 3(4), 364-377.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- 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 & Organization*, 34(2), 211-238.
- Wang, L. & Murnighan, J.K. (2013). Trust, White Lies, and Harsh Truths. Working Paper. City University of Hong Kong.
- Weber, J. M., Malhotra, D., & Murnighan, J. K. (2004). Normal acts of irrational trust:

Motivated attributions and the trust development process. *Research in Organizational Behavior*, 26, 75-101.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115(2), 157-168.

APPENDIX

Appendix A. Materials for Rely-or-Verify (Study 4)

Instructions to participants in Study 4:

As **BLUE Player**, you will report whether the amount of money in the jar of coins below is ODD or EVEN:



If you correctly report whether the amount of money in the jar is ODD or EVEN, you will earn \$1.50. If you inaccurately report if the amount of money in the jar is ODD or EVEN, you will earn \$0.

Your partner is **RED Player**. **RED Player's** bonus payment also depends on your choice.

[page break]

You will be asked to report whether the amount of money in this jar of coins is ODD or EVEN.

1. Before you do, you will receive a message from **RED Player**.

- **RED Player** knows the amount of money in the jar.
- **RED Player** will send you one of the following two messages: "The amount of money in the jar is ODD." or "The amount of money in the jar is EVEN."
- The message that **RED Player** sends could be *accurate* or *inaccurate*.
- *You will not know whether or not the message is accurate or inaccurate when you make the choice to **RELY** on the message or **VERIFY** the message.*

2. Once you've received the message from **RED Player**, you can choose to **RELY** on **RED Player's** message, or you can **VERIFY** the message.

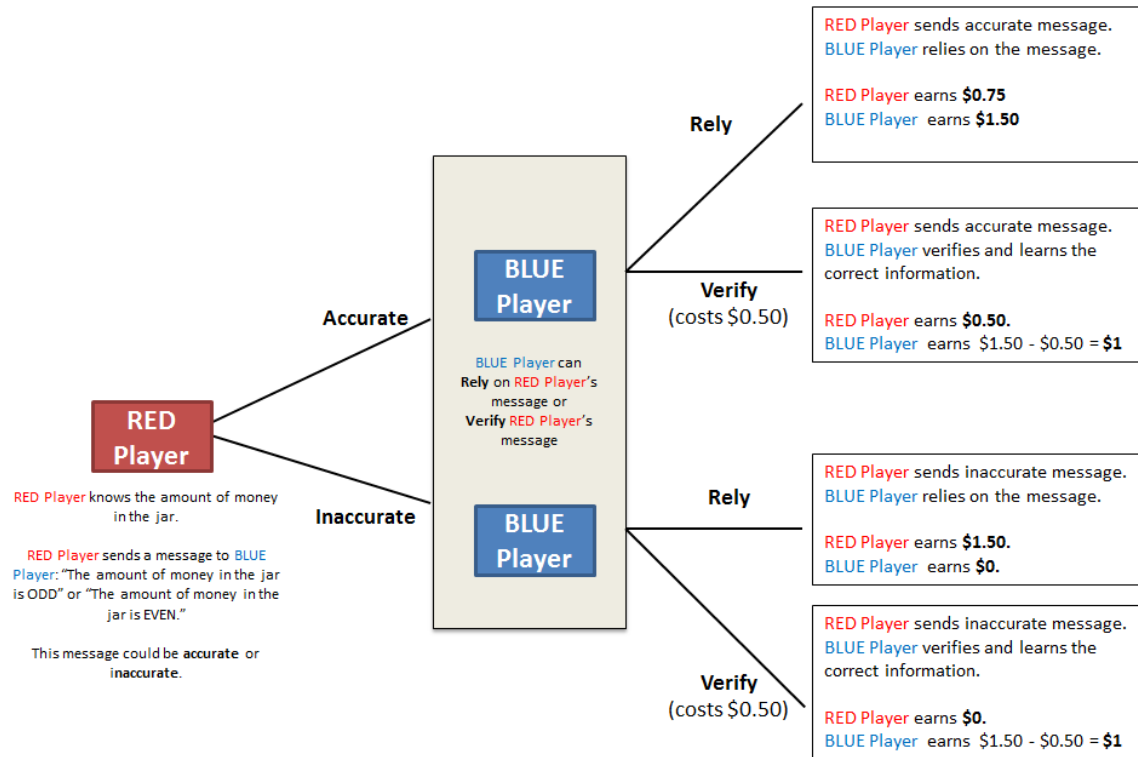
3. If you choose **RELY**, you will be paid based on whether or not **RED Player** gave you *accurate* or *inaccurate* information.

- If **RED Player** gave you *accurate* information and you **RELY** on it, you will correctly report whether the amount of money in the jar is ODD or EVEN, and **you** will earn \$1.50. **RED Player** will earn \$0.75.
- If **RED Player** gave you *inaccurate* information and you **RELY** on it, you will incorrectly report whether the amount of money in the jar is ODD or EVEN, and **you** will earn \$0. **RED Player** will earn \$1.50.

4. If you choose **VERIFY**, \$0.50 will be subtracted from your total earnings and you will learn the correct amount of money in the jar.

- If **RED Player** gave you *accurate* information and you **VERIFY** it, you will earn \$1 (\$1.50 for the correct answer - \$0.50 cost of verification) and **RED Player** will earn \$0.50.
- If **RED Player** gave you *inaccurate* information and you **VERIFY** it, you will earn \$1 (\$1.50 for the correct answer - \$0.50 cost of verification) and **RED Player** will earn \$0.

Your decisions are represented in the figure below.



Comprehension check questions for Rely-or-Verify:

1. Suppose **RED Player** sends you an *accurate message*. Will you earn more if you **RELY** or **VERIFY**?
2. Suppose **RED Player** sends you an *inaccurate message*. Will you earn more if you **RELY** or **VERIFY**?
3. How much does it cost to **VERIFY**?
4. If you **RELY** on **RED Player's** message, would **RED Player** earn more if s/he had sent a message that was *accurate* or *inaccurate*?

Appendix B. Items used to measure attitudinal trust in Trust game and Rely-or-Verify (Study 4)

- I trust my partner. [Rely-or-Verify uses identical measure].
- I am willing to make myself vulnerable to my partner. [Rely-or-Verify uses identical measure].
- I am confident that my partner will return half the money. [I am confident that my partner sent me an accurate message.]

Note. All items were anchored at 1 = “Strongly disagree” and 7 = “Strongly agree.”

Tables

Table 1. Payoffs associated with lying and honesty in Studies 1a, 1b, 2, 3a, and 3b

| | Experienced or Observed Deception | Deception Game | Type of Lie | | Payoffs associated with Truth (Option A) | Payoffs associated with Lie (Option B) |
|--|---|------------------------|----------------------------|----------|---|---|
| Study 1a | Experienced | Coin Flip Game | Altruistic Lie | Sender | \$2.00 | \$1.75 |
| | | | | Receiver | \$0.00 | \$1.00 |
| Study 1b | Observed | Coin Flip Game | Altruistic Lie | Sender | \$2.00 | \$1.75 |
| | | | | Receiver | \$0.00 | \$1.00 |
| Study 2 | Observed | Coin Flip Game | Prosocial Lie | Sender | \$2.00 | \$2.00 |
| | | | | Receiver | \$0.00 | \$1.00 |
| | | | Mutually beneficial Lie | Sender | \$2.00 | \$2.25 |
| | | | | Receiver | \$0.00 | \$1.00 |
| Studies 3a and 3b^a | Observed | Number Game (3a) | Altruistic Lie | Sender | \$2.00 | \$1.75 |
| | | | | Receiver | \$0.00 | \$1.00 |
| | | Coin Flip Game (3b) | Selfish Lie | Sender | \$1.75 | \$2.00 |
| | | | | Receiver | \$1.00 | \$0.00 |

Note. ^a Study 3b also included two control conditions. In control condition 1, the Sender faced the Altruistic Lie choice set, and in control condition 2, the Sender faced the Selfish lie choice set. However, in both control conditions, the Sender's decision was unknown.

Table 2. Descriptive statistics and correlations for measures in Studies 1, 2, and 3

| Study 1a | | | | | |
|----------------------|--------------------|--------------------|---------|---------|--|
| Scale | M(SD) | 1 | 2 | 3 | |
| 1. Trusting behavior | 43.8% ^a | | | | |
| 2. Attitudinal trust | 3.23 (1.91) | 0.88** | | | |
| 3. Benevolence | 3.82 (1.48) | 0.51** | 0.64** | | |
| 4. Deception | 4.10 (1.84) | 0.09 | 0.08 | -0.08 | |
| Study 1b | | | | | |
| Scale | M(SD) | 1 | 2 | 3 | |
| 1. Trusting behavior | 29.6% ^a | | | | |
| 2. Attitudinal trust | 3.08(1.65) | 0.70** | | | |
| 3. Benevolence | 3.95 (1.25) | 0.47** | 0.61** | | |
| 4. Deception | 4.15 (1.72) | -0.11 ⁺ | -0.13* | -0.29** | |
| Study 2 | | | | | |
| Scale | M(SD) | 1 | 2 | 3 | |
| 1. Trusting behavior | 50.2% ^a | | | | |
| 2. Attitudinal trust | 3.41(1.88) | 0.73** | | | |
| 3. Benevolence | 4.10 (1.33) | 0.49** | 0.63** | | |
| 4. Deception | 4.13 (1.86) | 0.08 | 0.01 | 0.05 | |
| Study 3a | | | | | |
| Scale | M(SD) | 1 | 2 | 3 | |
| 1. Trusting behavior | 36.2% ^a | | | | |
| 2. Attitudinal trust | 3.25(1.84) | 0.72** | | | |
| 3. Benevolence | 4.12 (1.40) | 0.41** | 0.67** | | |
| 4. Deception | 4.09 (2.42) | -0.12* | -0.25** | -0.34** | |
| Study 3b | | | | | |
| Scale | M(SD) | 1 | 2 | 3 | |
| 1. Trusting behavior | 47.2% ^a | | | | |
| 2. Attitudinal trust | 3.31(1.95) | 0.72** | | | |
| 3. Benevolence | 4.16 (1.40) | 0.68** | 0.68** | | |
| 4. Deception | 3.92 (2.27) | -0.26** | -0.26** | -0.38** | |

Notes. ^a This number represents the percent of participants who passed in the trust game.
 ** $p < .001$, * $p < .05$, ⁺ $p = .07$

Table 3. Supplemental regressions for Study 3a

| <i>Logistic regression on Trusting Behavior</i> | | | | |
|---|---|---|---|--|
| | (1) Intentions, Lying, Intentions x Lying | (2) Intentions, Lying, Intentions x Lying, Benevolence | (3) Intentions, Lying, Intentions x Lying, Deception | (4) Intentions, Lying, Intentions x Lying, Benevolence Deception |
| Constant | -.601** (0.122) | -3.887** (0.601) | 0.766 ⁺ (0.411) | -2.973*** (0.839) |
| Intentions | 0.498** (0.122) | 0.019 (0.151) | 0.482** (0.125) | 0.042 (0.153) |
| Lying | -0.002 (0.122) | 0.177 (0.134) | .697** (0.244) | .506 ⁺ (0.261) |
| Intentions x Lying | 0.032 (0.122) | -0.005 (0.131) | 0.025 (0.125) | -0.005 (0.132) |
| Benevolence | | 0.769** (0.133) | | 0.709*** (0.139) |
| Deception | | | -0.343** (0.100) | -0.166 (0.111) |
| R-Squared | 0.054 | 0.165 | 0.093 | 0.181 |

Notes. ***b* is significant at $p \leq .01$, * = $p < .05$. Standard errors are in parentheses.

Independent variables used in each regression are listed in the top row.

Deception was contrast-coded: -1 = Honest, 1 = Lie. *Intentions* was contrast-coded: -1 = Selfish, 1 = Prosocial.

Table 4. The payoffs associated with prosocial lying in Study 4

| | Type of Lie | | Payoffs associated with Truth | Payoffs associated with Lie |
|---------|-----------------------------|----------|----------------------------------|--------------------------------|
| Round 1 | Altruistic Lie | Sender | \$2.00 | \$1.50 |
| | | Receiver | \$0.25 | \$1.00 |
| Round 2 | Mutually- beneficial Lie | Sender | \$1.50 | \$2.00 |
| | | Receiver | \$0.25 | \$1.00 |
| Round 3 | Altruistic Lie | Sender | \$1.25 | \$1.00 |
| | | Receiver | \$0.25 | \$1.00 |
| Round 4 | Mutually- beneficial Lie | Sender | \$1.00 | \$1.25 |
| | | Receiver | \$0.25 | \$1.00 |

Table 5. Descriptive statistics and correlations for measures in Study 4

| Trust Game | | | | |
|-----------------------|---------------------|---------|---------|---------|
| Scale | M(SD) | 1 | 2 | 3 |
| 1. Trusting behavior | 48.50% ^a | | | |
| 2. Attitudinal trust | 3.82 (1.99) | 0.84** | | |
| 3. Benevolence | 4.44 (1.55) | 0.49** | 0.70** | |
| 4. Deception | 3.83 (2.34) | 0.07 | 0.03 | 0.06 |
| Rely-or-Verify | | | | |
| Scale | M(SD) | 1 | 2 | 3 |
| 1. Trusting behavior | 47.20% ^b | | | |
| 2. Attitudinal trust | 4.01 (1.73) | 0.80** | | |
| 3. Benevolence | 4.38 (1.42) | 0.41** | 0.65** | |
| 4. Deception | 3.76 (2.31) | -0.25** | -0.39** | -0.21** |

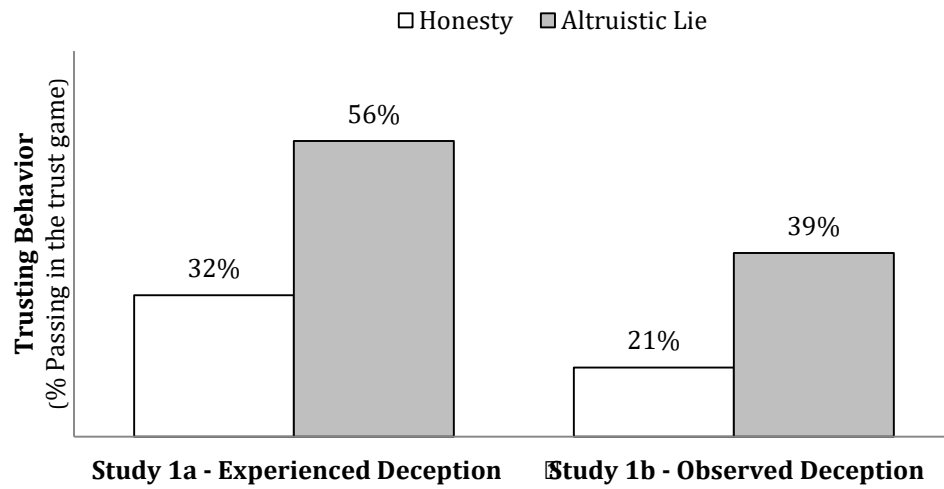
Notes. ** $p < .001$.

^a This number represents the percent of participants who passed in the trust game

^b This number represents the percent of participants who chose “Rely” in Rely-or-Verify.

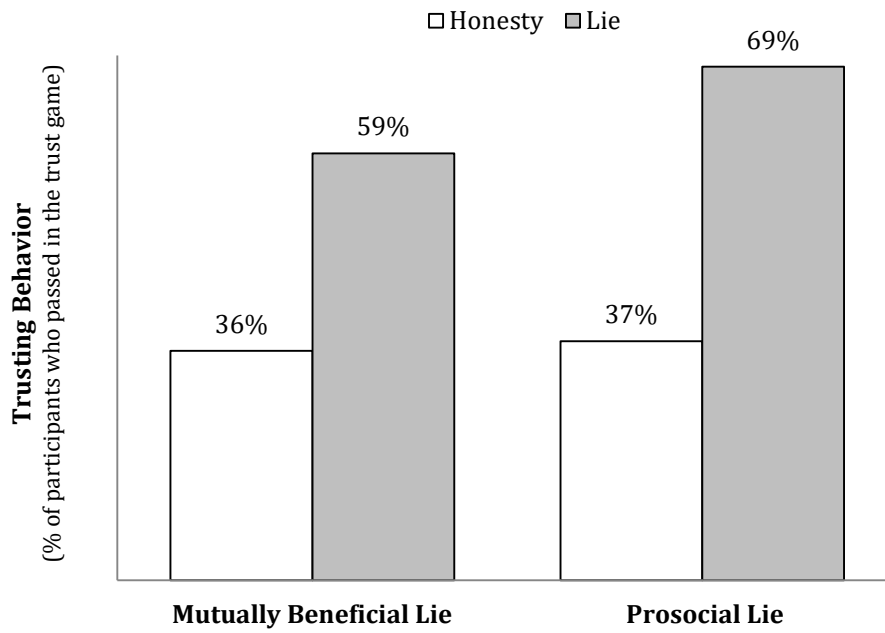
Figures

Figure 1. The effect of altruistic lying on trusting behavior (Studies 1a and 1b).



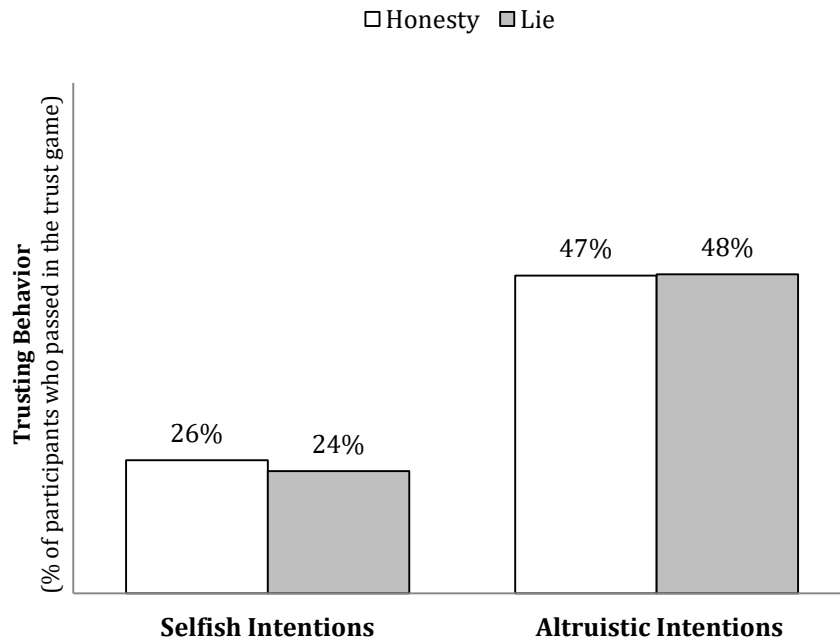
Note. Main effect of altruistic lying in both studies: $ps < .01$.

Figure 2. The effect of prosocial and mutually beneficial lying on trusting behavior (Study 2)



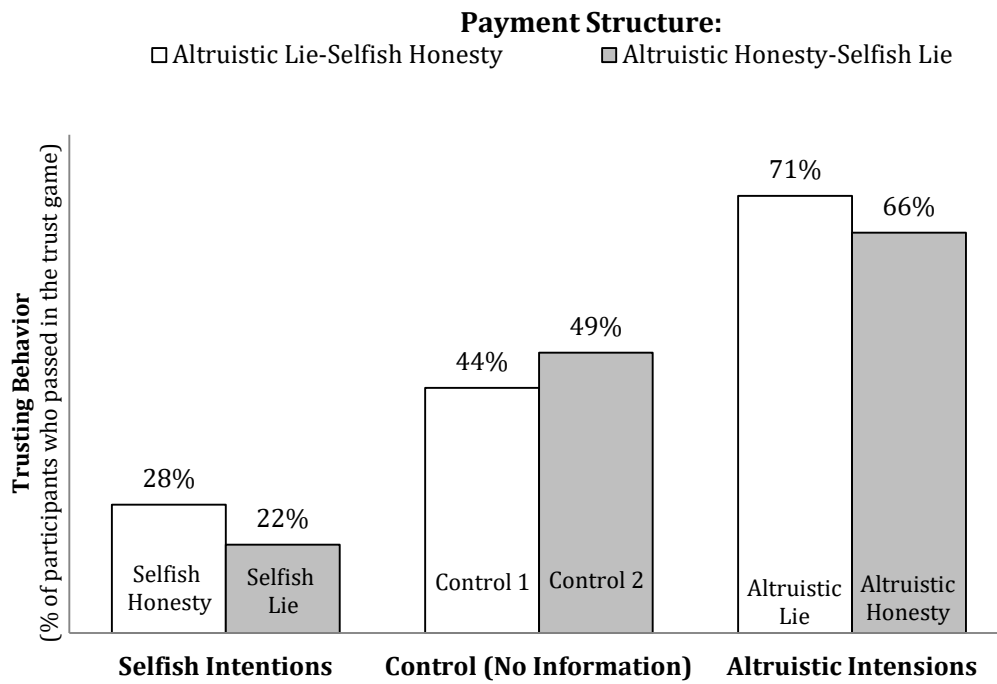
Note. Effect of lying for mutually-beneficial and prosocial lies: each $p < .01$.

Figure 3. Trusting behavior (Study 3a).



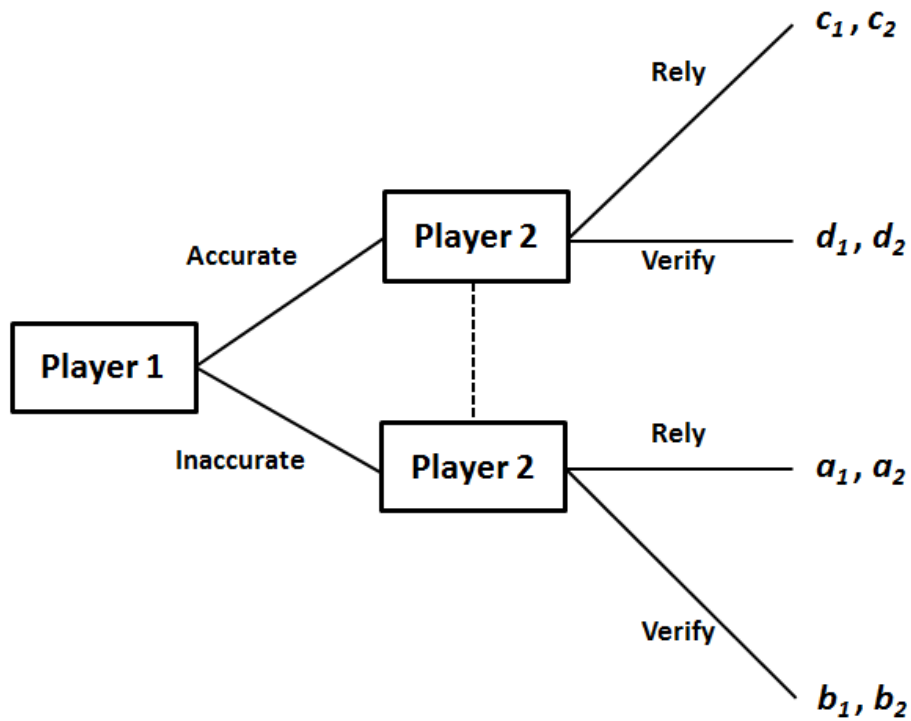
Note. Main effect of intentions: $p < .01$. Main effect of lying: *ns*.

Figure 4. Trusting behavior (Study 3b).



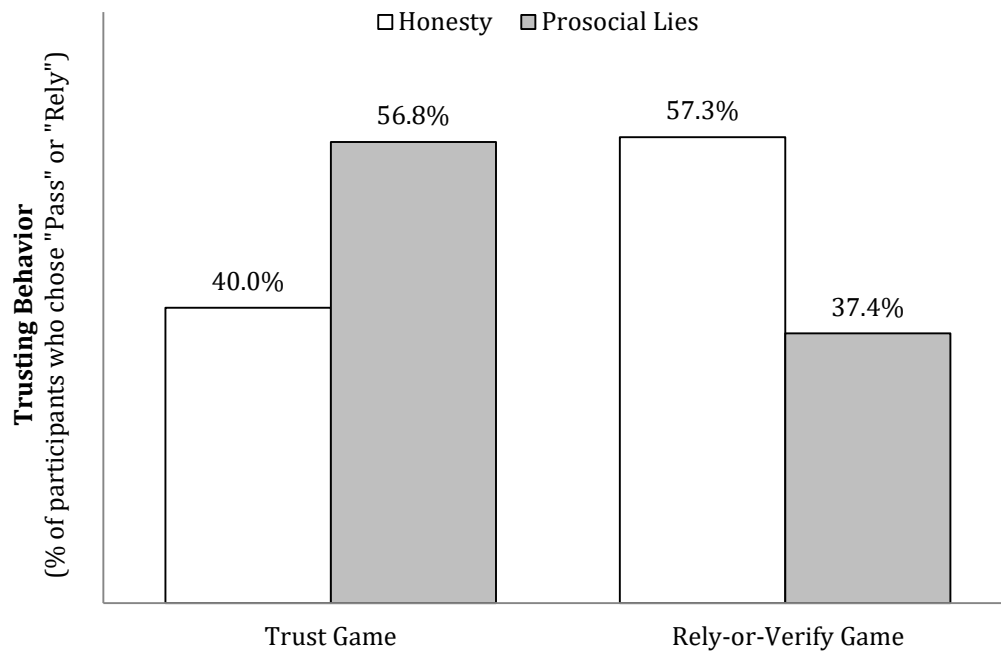
Note. Main effect of decision (Selfish, Control, Altruistic): $p < .01$. Main effect of payment structure: *ns*.

Figure 5. The Rely-or-Verify game (Study 4)



Note. This depicts the general form of Rely-or-Verify. The exact game we used in Study 4 is depicted in Appendix A. In Rely-or-Verify, the payoffs for Player 1 are structured such that $a_1 > c_1 \geq d_1 > b_1$. The payoffs for Player 2 are structured such that $c_2 > d_2 \geq b_2 > a_2$.

Figure 6. Trusting behavior (Study 4).



Note. Deception x Game interaction: $p < .01$. Main effects of deception and game: *ns*.