

NORTHWESTERN UNIVERSITY

Punishing deception and rewarding honesty

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

For the degree

DOCTOR OF PHILOSOPHY

Field of Management and Organizations

By

Cynthia S. Wang

EVANSTON, ILLINOIS

DECEMBER 2007

ABSTRACT

Punishing deception and rewarding honesty

Cynthia S. Wang

Encouraging honest behavior and discouraging deceptive behavior has become a very important issue with respect to ethics and moral behavior in the workplace. As such, a critical practical and theoretical goal has been to directly study the likelihood of individuals punishing deception and rewarding honesty. The growing assumption within the economic (Abbink, Irlenbusch, & Renner, 2000; Brandts & Charness, 2003; cf. Fehr & Gächter, 2000b; Offerman, 2002) and psychology literatures (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Cacioppo & Gardner, 1999; Rozin & Royzman, 2001) is that deception is punished more than honesty is rewarded. In contrast, my dissertation, drawing a theoretical foundation from the extant literature on the norm of reciprocity (Gouldner, 1960) and felt obligation (Cialdini, 1984; Eisenberger, Huntington, Hutchison, & Sowa, 1986; Pillutla, Malhotra, & Murnighan, 2003; Tesser, Gatewood, & Driver, 1968), suggests that honesty may be rewarded more than deception is punished. This paper investigates responses to deception and honesty in three ways: 1) comparing how deception is punished versus how honesty is rewarded in terms of frequency and intensity, 2) understanding how two fundamental factors, monetary costs and cultural influences, affect the frequency and intensity of punishments and rewards, and 3) understanding subsequent unethical behavior in an unrelated context after experiencing deception or honesty. I also propose a model outlining the underlying psychological processes motivating responses to deception and honesty in interpersonal situations. Results indicate that the punishment of deception and the reward of honesty arise from two distinct psychological mechanisms, with

negative emotions driving punishments and trust driving both felt obligation to reciprocate and rewards. The implications for social psychological, economic, and organizational theories and the practical implications for developing policies that promote ethical behavior within the workplace are discussed.

ACKNOWLEDGEMENTS

I am eternally indebted to a number of individuals who have helped me complete my dissertation. I first must thank my dissertation committee, David Austen-Smith, Adam Galinsky, Keith Murnighan, Katherine Phillips, for providing me with invaluable advice and direction. I especially want to thank Keith Murnighan for his incomparable mentorship, patience, and keen insight. It was an honor to have him as my dissertation chair.

I benefited from ample resources provided to me by the Dispute Resolution Research Center at Kellogg. The DRRC funded my dissertation and provided an open and lively venue for me to discuss my research ideas. This project also benefited from the research assistance of many individuals, including Iris Chae, Laura Howland, John Leahey, Marc Rhee, and Will Thomsen.

I also want to thank those who took time out of their busy lives to help improve my dissertation. Sharmi Surianarain, Deborah-Anna Reznik and Jennifer Whitson used their discerning eyes to provide feedback and make my dissertation comprehensible to most. Mark Rivera and Chris Bauman helped me with their quantitative prowess. My dissertation would have gone astray without the theoretical insight and advice from Niro Sivanathan. I owe Adam Galinsky my gratitude; our discussions were an essential source of inspiration for my dissertation.

Of course, I could not have completed this dissertation without the tremendous emotional support from my parents, Ko-Hsueh and Andrew Wang, my brothers, Alex and Yuan Wang, and those I consider family, Ben Geerts, Jean Hong, and Kodzai Nyakurimwa and Deborah-Anna Reznik.

TABLE OF CONTENTS

ABSTRACT	3
ACKNOWLEDGEMENTS.....	5
LIST OF TABLES.....	7
LIST OF FIGURES	8
INTRODUCTION	10
CHAPTER 1. Experiment 1: Punishing deception versus rewarding honesty	26
CHAPTER 2. Experiment 2: Why recipients reward more than they punish	34
CHAPTER 3. Cultural differences in rewarding honesty and punishing deception	58
CHAPTER 4. Experiment 3: Cultural differences in scenario responses.....	63
CHAPTER 5. Experiment 2 (reanalyzed by culture)	68
CHAPTER 6. General Discussion.....	74
TABLES	98
FIGURES.....	106
APPENDICES	128

LIST OF TABLES

<i>Table 1.</i> Gender frequencies by experimental condition, Experiment 1.....	98
<i>Table 2.</i> Gender frequencies by experimental condition, Experiment 2.....	99
<i>Table 3.</i> Means, standard deviations, and correlations for independent and dependent variables, Experiment 2.....	100
<i>Table 4.</i> Regression estimations for trust and negative affect moderated mediation analyses (intensity), Experiment 2.	102
<i>Table 5.</i> Logistic regression estimations for trust and negative affect moderated mediation analyses (pen theft), Experiment 2.....	103
<i>Table 6.</i> Means, standard deviations, and correlations for independent and dependent variables, Experiment 2 reanalyzed.....	104

LIST OF FIGURES

<i>Figure 1.</i> Model outlining responses to deception versus honesty	106
<i>Figure 2.</i> The effect of partner's behavior and cost on choice to respond (percentages and frequencies), Expt 1.	107
<i>Figure 3.</i> The effect of partner's behavior and cost on amount of response, including zero responses (Means \pm SEM and frequencies per condition), Expt 1	108
<i>Figure 4.</i> The effect of partner's behavior and cost on amount of response, excluding zero responses (Means \pm SEM and frequencies per condition), Expt 1	109
<i>Figure 5.</i> The effect of the other player's behavior and cost on choice to respond (percentages and frequencies), Expt 2.	110
<i>Figure 6.</i> The effect of the other player's behavior and cost on intensity of response, including zero responses (Means \pm SEM and frequencies per condition), Expt 2.	111
<i>Figure 7.</i> The effect of the other player's behavior and cost on intensity of response, excluding zero responses	112
<i>Figure 8.</i> Trust and negative affect moderated mediation analyses (intensity), Expt 2.	113
<i>Figure 9.</i> The effect of trust on intensity, Expt 2.....	114
<i>Figure 10.</i> The effect of negative affect on intensity, Expt 2.	115
<i>Figure 11.</i> Felt obligation mediation analysis: Felt obligation as the mediator between trust and reward, Expt 2.....	116
<i>Figure 12.</i> The effect of the other player's behavior and cost on pen theft, Expt 2.	117
<i>Figure 13.</i> Trust and negative affect moderated mediation analysis (pen theft), Expt 2.....	118

<i>Figure 14.</i> The effect of partner’s behavior and culture on choice to respond (percentages and frequencies), Expt 3.	119
<i>Figure 15.</i> The effect of partner’s behavior and culture on amount of response, including zero responses (Means \pm SEM and frequencies per condition), Expt 3.	120
<i>Figure 16.</i> The effect of partner’s behavior and culture on amount of response, excluding zero responses (Means \pm SEM and frequencies per condition), Expt 3.	121
<i>Figure 17.</i> The effect of partner’s behavior and culture on frequency of response (percentages and frequencies), Expt 2 reanalyzed.	122
<i>Figure 18.</i> The effect of partner’s behavior and culture on intensity of response, including zero responses (Means \pm SEM and frequencies per condition), Expt 2 reanalyzed.	123
<i>Figure 19.</i> The effect of partner’s behavior and culture on intensity of response, excluding zero responses (Means \pm SEM and frequencies per condition), Expt 2 reanalyzed.	124
<i>Figure 20.</i> The effect of partner’s behavior and culture on trust, Expt 2 reanalyzed.	125
<i>Figure 21.</i> The effect of partner’s behavior and culture on unhappiness, Expt 2 reanalyzed. ...	126
<i>Figure 22.</i> The effect of partner’s behavior and culture on felt obligation, Expt 2 reanalyzed.	127

INTRODUCTION

Knowing how to encourage positive behavior and deter negative behavior within an organizational setting has become an important line of inquiry for managers. The recent accounting scandals plaguing corporations within the past decade have made this topic even more salient and relevant. Understanding when and why people punish deceptive behavior and reward honest behavior is critical in creating policies that promote desirable and ethical behaviors within organizations.

As a result of the growing practical importance of this topic, studying responses to negative and positive behaviors such as deception and honesty has received ample attention in the fields of organizational behavior (Cropanzano & Mitchell, 2005), economics (Fehr & Gächter, 2000b; Fehr & Schmidt, 1999) and social psychology (Eisenberger, Lynch, Aselage, & Rohdieck, 2004; Gouldner, 1960). More recently, economic researchers have become increasingly interested in comparing and contrasting frequencies and intensities of punishment versus reward, with the prevailing assumption that deception is punished more than honesty is rewarded (Abbink et al., 2000; Brandts & Charness, 2003; cf. Fehr & Gächter, 2000b; Offerman, 2002).

However, the economic research on rewarding honesty and punishing deception, still in a nascent stage, displays some limitations, e.g. in the study designs, levels of deception and honesty are not equal in magnitude. These asymmetries in design leave the question open to whether deception is punished more than honesty is rewarded or whether honesty is rewarded more than deception is punished. The current research attempts to clarify how and why deception is punished and honesty is rewarded, along with conjecturing that, contrary to the

recent assumptions, recipients may actually reward honesty *more* than they punish deception, with the asymmetry in response driven by possible asymmetric reciprocity norms and feelings of obligation (Cropanzano & Mitchell, 2005; Eisenberger et al., 2004).

In four experiments, I investigate reactions to deception and honesty, attempting to uncover the underlying mediating and moderating mechanisms. In the experiments, I refer to the deceptive or honest actor as ‘the actor’ and the recipient of the deceptive or honest act as ‘the recipient’, with particular interest in the behavior of the recipient. I studied *direct responses* to deceptive and honest behavior, specifically the punishment and reward of actors by recipients. Furthermore, I explored how two fundamental factors, monetary costs to respond and cultural norms, alter the frequency and intensity of punishment and reward. In addition to measuring direct responses, I also measured *indirect responses* to deception and honesty. In particular, in contrast to studying the responses of recipients toward the actors themselves, I studied their subsequent deviant behavior (e.g., theft) in an unrelated context following actors’ deception or honesty. Studying indirect responses is ultimately important in assessing whether deception and honesty are consequential for not only direct reciprocal exchanges, but also for the unrelated surrounding situational context.

In addition to examining the behavioral responses, another significant undertaking of this research was to understand the affective and cognitive mechanisms driving both the direct and indirect responses to deception and honesty. Overall, this paper attempts to provide a theoretical model that outlines individuals’ responses to deception and honesty (see Figure 1), with the experiments improving upon past research by providing a clear comparison of punishment and reward given equivalent levels of deception and honesty. The next section reviews the

reciprocity literature within the economics field, which gives insight into how recipients punish actors' negative behavior (e.g., deception) and reward actors' positive behavior (e.g., honesty).

Negative and Positive Reciprocity

Recent research in behavioral game theory has investigated conditions under which recipients exhibit negative reciprocity, i.e. punishing actors who treat them unfavorably (Boles, Croson, & Murnighan, 2000; Kahneman, Knetsch, & Thaler, 1986; Pillutla & Murnighan, 1996), along with positive reciprocity, i.e. rewarding actors who treat them favorably (Berg, Dickhaut, & McCabe, 1995; McCabe, Rassenti, & Smith, 1998). The practice of reciprocity departs from game theoretic predictions of subgame perfect equilibria (Selten, 1965). Because any amount offered by the proposer will be monetarily beneficial to the responder, the subgame perfect equilibrium requires the responder to accept any positive offer (Selten, 1965). Recent experiments show that recipients reciprocate even at their own monetary expense (Boles et al., 2000; Kahneman et al., 1986; Pillutla & Murnighan, 1996) and when no material gains are expected from the interactions (e.g., one-shot interactions, see Fehr & Gächter, 2000b, for a review).

Two seminal economic games that illustrate negative reciprocity are the ultimatum game and the public goods game with punishment. In an ultimatum game, a proposer can offer any amount of money to a responder. The responder then must decide whether to accept or to reject the offer. Studies have consistently found that, with high probability, responders tend to reject offers that are less than 30 percent of the proposers' allocations (see Camerer & Thaler, 1995; Guth, Schmittberger, & Schwarze, 1982; Kahneman et al., 1986, for examples).

The public goods game with punishment is another influential economic game illustrating negative reciprocity. Here, free riders are punished even at a personal cost to the punisher. In the public goods game, the dominant individual strategy is to free ride, but the most effective collective strategy is to contribute all assets to the public good. For example, in an experiment by Fehr & Gächter (2000a), four group members were given 20 tokens each and decided how many tokens to keep for themselves or to invest in a common public goods pool. A token kept by a group member did not change in value. However, for each token invested into the public goods pool by a group member, each group member received .4 tokens, a total social return of 1.6 tokens. Thus, the most effective collective strategy was for group members to contribute all assets to the public good because each group member received 32 tokens or the group as a whole received 128 tokens. However, the dominant individual strategy was to contribute zero to the public good because the individual group member did better by contributing zero than by contributing any other amount, regardless of what others contributed. In the punishment version of this game, group members could punish at a cost to themselves. If group members were purely self-interested, they would never partake in costly punishment. However, group members consistently punished free riders and, as a result, achieved very high levels of cooperation (Fehr & Gächter, 2000a; Ostrom, Gardner, & Walker, 1994).

Trust and gift exchange games also illustrate that positive reciprocity occurs, also contrary to subgame-perfect equilibrium predictions (Berg et al., 1995; Fehr, Kirchsteiger, & Riedl, 1993; McCabe, Rassenti, & Smith, 1996). Reciprocity entails not only hurting those who hurt us, but also helping those who help us. In a trust game, a proposer can offer any amount of money to the responder. The experimenter then triples the amount of money sent (with the

knowledge of both players), so that the responder receives three times amount sent. The responder then chooses any amount up to the total tripled amount to return to the proposer. A substantial number of responders return more than the amount sent by proposers, even though returning zero maximizes the responders' monetary outcomes (Berg, Dickhaut, McCabe, 1995).

Evidence of positive reciprocity also exists in non-experimental settings. In a classic study of positive reciprocity, researchers from Utah sent 500 Christmas cards to strangers in Chicago, 20% of whom reciprocated by returning cards or letters without question or concern about the unknown identities of the senders (Kunz & Woolcott, 1976).

These studies demonstrate both negative and positive reciprocity at a cost. While these studies do not directly compare levels of positive and negative reciprocity, more recent explorations have begun to examine this topic (Abbink et al., 2000; Brandts & Charness, 2003; Offerman, 2002). In the next section, I review the recent theoretical discussions directly comparing a variety of negative and positive phenomena (Baumeister et al., 2001; Cacioppo & Gardner, 1999; Rozin & Royzman, 2001), along with the experimental economics literature specifically comparing negative and positive reciprocity. Both literatures posit stronger responses to negative events than to positive events.

Punishing deception > Rewarding honesty

In the past thirty years, an abundance of research has focused on the dismal decision-making processes of individuals and groups (see Wang & Thompson, 2006, for a review), incorrectly labeling individuals as cognitive misers (S. T. Fiske & Taylor, 1991), biased decision-makers (Tversky & Kahneman, 1974), and faulty negotiators (Bazerman, Magliozzi, & Neale, 1985).

Within the last decade, research has also involved the ‘positive psychology’ of human behavior, the study of positive individual and institutional features (Seligman & Csikszentmihalyi, 2000). More recently, comparing responses to good versus bad stimuli have become a central theoretical topic in the fields of social psychology (see Baumeister et al., 2001; Rozin & Royzman, 2001, for a review), organizational behavior (see Wang & Thompson, 2006, for a review), and economics (Falk & Fischbacher, 2006; Fehr & Gächter, 2000b), focusing on understanding negative and positive phenomena simultaneously.

The greater power of the negative over the positive in almost every aspect of life is outlined in a number of comprehensive social psychology literature reviews (Baumeister et al., 2001; Cacioppo & Gardner, 1999; Rozin & Royzman, 2001). Rozin and Royzman (2001) outline a strong negativity bias: given the same magnitude, negative events, compared to positive events, seem to affect emotions, cognitions, and behavior to a greater extent.

The greater impact of the negative is apparently universal, ranging from close relationship outcomes to learning and social support (Baumeister et al., 2001). For example, one strongly supported finding is the positive-negative asymmetry effect: the propensity for individuals to process negative information more thoroughly than positive information, and, in turn, the tendency to use the negative information to form final impressions (e.g., Anderson, 1965; Peeters & Czapinski, 1990; Skowronski & Carlston, 1989).

There is evidence that perceptions and judgments of actors are swayed more by their negative behaviors than by their positive behaviors. Actors who behave negatively, compared to those who behave positively, are viewed as stronger within a prisoner’s dilemma (Exline & Baumeister, 2007, as cited in Baumeister, et al., 2001) and are gazed at longer in photographs (S.

T. Fiske, 1980, as cited in Baumeister, et al., 2001). Negative acts also lead to more dislike than positive acts lead to like (Falk & Fischbacher, 2006; S. T. Fiske, 1980, as cited in Baumeister, et al., 2001).

While the psychological reviews provide evidence that perceptions are influenced more by negative behaviors than by positive behaviors, the assumption within the reviews that recipients reciprocate negative behaviors more than positive behaviors is only implicit. Behavioral economists have focused more on directly comparing negative and positive reciprocity, comparing and contrasting how recipients punish negative behavior and reward positive behavior. Several recent economic models address both positive and negative reciprocity (e.g., Dufwenberg & Kirchsteiger, 2000; Fehr & Schmidt, 1999; Rabin, 1993). Whereas these models include some implicit assumptions about the asymmetries of positive and negative reciprocity, they do not explicitly differentiate the strengths of these two acts. For example, Rabin (1993) discusses how both negative and positive reciprocity can occur, but does not explicitly delineate how similar levels of unkindness and kindness might result in dissimilar levels of punishment and reward.

More recent economic experiments measure and compare positive and negative reciprocity, with the emerging assumption that harmful behavior is punished more intensely than friendly behavior is rewarded (cf. Fehr & Gächter, 2000b). For example, in the ultimatum game, a proposer offering nothing is rated more harshly than a proposer offering everything is rated favorably (Falk & Fischbacher, 2006). Recent findings show that negative behaviors are reciprocated more often than positive behaviors (Abbink et al., 2000; Brandts & Charness, 2003; Engelmann & Ortmann, 2001; Offerman, 2002; Pereira, Silva, & Silva, 2006).

For example, Brandts and Charness (2003) examined reactions to misleading messages in one-shot environments. In Stage 1 of their study, an actor sends a costless and non-binding message (cheap talk) to a recipient, stating his or her intended move. One move produces higher payoffs for the receiver (6 units for the actor, 9 units for the recipient), the other move produces higher payoffs for the sender (12 units for the actor, 3 units for the recipient). In Stage 2, the actor can choose to follow through on the message or choose the other option. In Stage 3, the recipient can then redistribute payoffs as a reaction to the actor's actual move: the recipient can punish the sender for choosing the (12, 3) option by reducing the actor's payoff by 10 units and reducing his own payoff by 1 unit, leaving both the actor and recipient with 2 units.

Alternatively, the recipient can reward the actor with 2 units for choosing the (6, 9) option while incurring a cost of 2 units for the move, thereby leaving the actor with 8 units and the recipient with 7 units. Thus, recipients of deception inflict a higher monetary punishment than recipients of honesty bestow monetary reward. The results showed that 49% of those receiving an inaccurate message punished, whereas only 25% of those receiving an accurate message rewarded. The results support the argument that recipients punish deception more than reward honesty. However, the next section outlines key limitations that leave the question open to whether recipients punish or reward more.

Research limitations

Some of the experiments directly comparing frequencies of reward and punishment (e.g., Brandts & Charness, 2003; Offerman, 2002) exhibit notable asymmetries in the designs. For example, Brandts & Charness (2003), the payoff difference between actors and recipients was larger when the payoff was higher for actors (9 units difference) than when the payoff was higher

for recipients (3 units difference). Therefore, the higher frequency of punishment may have directly resulted from a bigger payoff difference following deception than following honesty. Another design asymmetry was that the ratio of costs to punishments was less than the ratio of costs to reward: for every 1 unit spent, recipients could punish 10 units. On the other hand, for every 1 unit spent, the recipient could only reward 1 unit.

Offerman (2002) discovered that 67% of recipients punished negative behavior and 25% rewarded positive behavior. This study displays a similar confound as Brandts & Charness (2003), with a larger payoff differential following negative actions (recipients receive 15 fewer Dutch guilders than actors) than following positive actions (recipients receive 4 more Dutch guilders than actors). In these studies, a lower cost to reward than punish and a greater comparative loss than gain both could bias recipients to reward honesty more than punish deception. Thus, the limitations within these studies provoke questions as to whether deception or honesty elicits stronger reactions.

Although methodologically imperfect evidence seems to suggest that recipients of deception punish more than recipients of honesty reward, the research on reciprocal norms (Gouldner, 1960) suggests that recipients of honesty may reward more than recipients of deception punish. Researchers have discussed the strength of reciprocal norms, of which the positive has been directly tied with feelings of indebtedness toward the benefactor and feelings of obligation to return bestowed benefits (Cialdini, 1984; Eisenberger et al., 1986; Pillutla et al., 2003; Tesser et al., 1968). The next section discusses literature that is consistent with the hypothesis that recipients might actually reward honesty more than they punish deception.

Rewarding honesty > Punishing deception

Reciprocal norms

The norm of reciprocity (Gouldner, 1960) asserts that one's actions should be paid back in kind. This concept is best captured in a number of well-known, historical phrases. "An eye for an eye" is a quotation originating from Exodus 21:23-25 expressing the principle of negative reciprocity, which suggests that hurtful treatment should be paid back in kind. In contrast, the notion to reciprocate positive treatment dates back to Ancient Rome when philosopher Marcus Cicero stated that "there is no duty more indispensable than that of returning a kindness" (cited in Gouldner, 1960, p. 161). A norm represents a standard that *ought* to be followed (e.g., Moore, 1903). Key in this concept is that interactions between an actor and a recipient generate feelings of obligation to give in return in an equivalent manner what one has received.

The positive reciprocity norm has received ample theoretical attention by social scientists. Within an organizational setting, Blau (1964) has posited that workers and employers strive to maintain a fair balance in their reciprocal exchanges. In his seminal article, Gouldner (1960) surmised that the norm was a universal principal and one of the most pervasive social forces in all human cultures. Gouldner's discussion, while focused primarily on the positive norm of reciprocity, also discussed the negative norm of reciprocity in terms of the belief that one should take retribution for inflicted bad behavior. Other authors have also discussed reprisal for abuse being encouraged by a negative reciprocity norm (Cialdini, Green, & Rusch, 1992; Helm, Bonoma, & Tedeschi, 1972; Youngs, 1986).

Recent explorations of the norms of reciprocity have included both positive and negative forms in their analyses, finding that the two types of reciprocities are not highly correlated and arise from distinct belief systems (Eisenberger et al., 2004; Uhl-Bien & Maslyn, 2003). For

example, endorsement of a negative reciprocity norm was related to a tendency to be angry in everyday life and to having beliefs that people are generally malicious, whereas the endorsement of a positive reciprocity norm was related to beliefs that people are generally kind and trustworthy (Eisenberger et al., 2004).

Felt obligation

One rationale for why recipients may reward honesty more than they punish deception is that the two norms of reciprocity may induce different levels of felt obligation. A plethora of evidence ties felt obligation to positive reciprocity. When one party's contributions generate an imbalance in the relationship, the other party feels obligated to reciprocate in a desire to regain balance (Eisenberger, Armeli, Rexwinkel, Lynch, & Rhoades, 2001; Eisenberger et al., 1986; Greenberg, 1980; Gross & Latane, 1974; Pillutla et al., 2003). Recent evidence supports this theory, with feelings of obligation driving the reciprocation of gifts, favors, and help (Cialdini, 1984; Eisenberger et al., 1986; Pillutla et al., 2003; Tesser et al., 1968). In contrast, there is little empirical evidence that directly ties negative reciprocity to felt obligation, suggesting that recipients may not feel as obligated to punish deception as they do to reward honesty.

The different levels of felt obligation may result from competing motivations, with honesty engendering a motivation to approach an honest actor and deception producing a motivation to avoid a deceiver. This thesis is drawn from the approach-avoidance motivation literature, which posits that approach behavior is instigated by a desirable event, whereas avoidant behavior is instigated by a negative event (Elliot, 1999; W. James, 1890). As a whole, a positive action can engender a desire to approach and cooperate with the target, thereby facilitating the development of relationships (Gross & Latane, 1974; Staub, 1972).

According to Hokanson's (1974) escape-avoidance model of aggression, individuals behave in a manner that reduces physiological discomfort arising from negative stimuli. The extent to which recipients have found retaliatory or non-retaliatory responses successful for escaping mistreatment influences whether they will endorse a negative norm of reciprocity (Eisenberger et al., 2004). Individuals who have learned to cope using retaliatory methods in the past would be more likely to feel obligated and act on the negative norm of reciprocity. Individuals who have used non-retaliatory responses may feel less obligated to respond and instead proffer a strong norm of avoidance and a desire to escape the situation. Moreover, especially in work situations where negatively reciprocating with bosses can have extremely detrimental consequences, there may be a greater norm to avoid rather than to confront (Duffy, Ganster, & Pagon, 2002; Tepper, 2000). Similarly within the group identity literature, most group bias consists of helping the ingroup while avoiding the outgroup (Brewer, 1999). Thus, for disliked individuals, instead of aggression and confrontation, many times the rationale seems to be more of avoidance.

The discussions about the strong obligations to positively reciprocate (Gouldner, 1960; Kunz & Woolcott, 1976) and the asymmetries in the economic research designs (Abbink et al., 2000; Brandts & Charness, 2003; Offerman, 2002) suggest that recipients may reward honesty more than they punish deception. Thus, the foremost goal of this paper is to compare responses to deception and honesty more precisely.

Monetary costs to respond

Impact of costs on rewards/punishments

Economic experiments often include monetary costs to reciprocate, namely to show that recipients do not always attempt to maximize their monetary outcome as predicted by game theory (cf. Fehr & Gächter, 2000b). More recent explorations vary levels of costs to punish. As one might expect, as costs to punish rise, recipients are less willing to reciprocate negatively (Brosig, Weimann, & Yang, 2003; Cameron, 1999; Leventhal & Anderson, 1970). For example, in a simulated ultimatum game, responders were more likely to punish (i.e., reject the offer) a 90% split when the amount to be split was \$1 than when it was \$1 million (Leventhal & Anderson, 1970). Similarly, in an actual ultimatum game run in Indonesia, responders were more willing to reject a given percentage offer when the monetary stakes were approximately .75 months income than when they were approximately 3 months income (Cameron, 1999).

In contrast, recipients continued to reward at similar proportions regardless of stake level (Fehr, Tougareva, & Fischbacher, 2002). Subjects from Russia played a gift exchange game where wages are exchanged for effort. Game outcomes determined the actual amounts subjects earned in the experiment. The same wage and effort patterns prevailed regardless of whether subjects earned between two and three times their monthly income or one week of their monthly income. This suggests that increasing costs might impact decisions to punish more than they do decisions to reward. While suggestive, the evidence does not directly compare punishments and rewards using equivalent cost structures (e.g., Brandts & Charness, 2003; Brosig et al., 2003). As a result, a central area of inquiry in this paper is to understand the moderating influence of cost on responses, directly testing how varying levels of cost impact rewards and punishments.

Loss Aversion

A well-established finding is that judgments differ depending on whether an outcome is viewed as a loss or a gain (Kahneman & Tversky, 1979). According to prospect theory, objective outcomes are compared to a reference outcome. When the objective outcome is less favorable than the reference outcome, the decision-maker is placed in a loss frame. When the objective outcome is more favorable than the reference outcome, the decision-maker views the situation in a gain frame. Deception and honesty can often be interpreted within a loss/gain frame, as deception often entails losses and honesty entails gains. Prospect theory posits a curvilinear relationship between the objective outcome and subjective utility, where the slope of the value function is steeper during losses than during gains, as a result, decision-makers prefer avoiding losses to acquiring gains.

This concept of loss aversion provides a basis for the argument that costs influence levels of punishment more than levels of reward. According to prospect theory, because the curve is steeper when in a loss frame (following deception) than when in a gain frame (following honesty), spending money to punish deception, compared to spending money to reward honesty, will result in a greater drop in subjective utility. In other words, spending money to punish will hurt more than spending money to reward and recipients will be less willing to spend money to punish deception than to reward honesty. This suggests that recipients will be less likely to punish than reward as costs increase.

Overview

The experiments in my dissertation eliminate three key asymmetries of past research (Brandts & Charness, 2003; Offerman, 2002). The first asymmetry was the larger payoff difference between actors and recipients following deception than following honesty. The

current experiments removed the social comparison of payoff differences by omitting information about the actors' payoffs. Thus, the recipients focused on the act of deception or honesty, rather than to their outcomes compared to the actors'. The second asymmetry was the higher cost to reward than to punish. The current experiments were designed so that the cost to punish and reward were equivalent per condition. The third asymmetry involved the levels of deception and honesty. Unlike past experiments where it was unclear whether the levels of deception and honesty were equivalent, deception and honesty were manipulated to have equivalent impact so that deception clearly hurt, in an objective sense, as much as honesty helped.

In four experiments, I provided a systematic comparison of how recipients reward honesty and punish deception in terms of frequency and intensity. To provide a clear comparison, I tested how recipients punished deception and rewarded honesty in the economic framework used in prior experiments, assessing how recipients responded when they never interacted with the actors again, did not expect any material gains following the interaction, and when it was costly to respond. Experiment 1 established that, given comparable levels of deception and honesty, recipients rewarded more than they punished. In Experiment 2, I examined the impact of felt obligation and other central mediating affective and cognitive variables to provide greater insight into the processes underlying responses to deception and honesty.

I also examine two important moderating variables that potentially alter levels of punishments and rewards: monetary costs and cultural differences. In the experiments, I systematically manipulated costs, making costs to punish and reward equal and symmetric. In

terms of cultural differences, evidence suggests that East Asians punish more intensely than Westerners within domains such as child rearing (Kelley & Tseng, 1992) and arbitrator behavior (Friedman, Liu, Chen, & Chi, 2007), and that these differences arise from different cultural norms of reciprocity (Chiu, Dweck, Tong, & Fu, 1997). However, there is less direct evidence comparing punishment and reward behavior in different cultures. As a result, in Experiments 3 and 2 (reanalyzed), I examined East Asian responses compared to Western responses.

CHAPTER 1

Experiment 1: Punishing deception versus rewarding honesty

The first study manipulated two factors – the actor’s behavior and the cost to respond. I manipulated the acts of deception and honesty to be equivalent and equal. I ensured that the change in absolute magnitude and relative magnitude of recipients’ outcomes as a result of deception and honesty were the same. In terms of absolute magnitude, the dollar amount gained after honesty was the same as the dollar amount lost after deception. For example, in Experiment 1, one scenario included an act of deception that resulted in the recipient receiving \$50 less than expected. The other scenario included an act of honesty that resulted in the recipient receiving \$50 more than expected; both acts resulted in a net change of \$50.

However, another view of equivalence is that the relative change in wealth of the decision maker should be equal (Thaler, 1980; Tversky & Kahneman, 1981). Referring again to the scenarios, the \$50 gain is not equivalent to the \$50 loss in terms of relative change. The rationale is as follows. Following deception, if recipients expected \$150 but then only received \$100, they would have perceived the \$50 loss in comparison to their expectation of \$150. Thus, the perceived loss is $50/150 \times 100\% =$ approximately a 33% decrease in perceived wealth. Following honesty, if recipients expected \$50 but received \$100, they would have perceived this as an increase of \$50 in relation to what they expected to get. In this case, the perceived increase is $50/50 \times 100\% = 100\%$, with the change representing a 100% increase in perceived wealth. To make an equivalent relative change of 33% following honesty, post-honesty recipients would have expected to receive \$75, rather than \$50. Thus, the perceived increase would have been $25/75 \times 100\% = 33\%$, with the change representing a 33% increase in perceived wealth. I

included both the relative- and absolute-honesty conditions and compared each to the deception condition.

Method

Participants

Participants were 107 undergraduate students (44 men, 60 women, 3 non-reporters) from a major Midwestern University. Participants filled out the survey as one part of a 45 minute multi-experiment session for which they were paid \$10.

Design and Procedures

Experiment 1 had a 3 (behavior of partner: deception, honesty-absolute, honesty-relative) x 2 (cost to respond: low, high) between-subjects design.¹ In Experiment 1, each participant read a scenario about a businessperson named Pat and filled out a response form. The six versions of the scenario can be found in Appendix 1.

The scenarios manipulated Pat's behavior and the cost to respond. Pat's behavior was manipulated in one of three ways: Pat either behaved dishonestly, honestly (equivalent to the dishonest act in an absolute manner), or honestly (equivalent to the dishonest act in a relative manner). Two of the manipulations are presented below, with the honest-absolute manipulation italicized and not parenthesized and the dishonest condition italicized and parenthesized:

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was *honest (dishonest)* about some key information regarding the deal. As a result, you (*only*) received \$100. You would have received 50% *less (more)* if Pat had given you *dishonest (honest)* information.

The scenarios controlled for wealth effects: recipients received \$100 regardless of their partners' behavior. In the dishonest condition, recipients received 50% less than they would have received if their partners had been honest. Thus, in absolute terms, recipients received 50% less than they would have received otherwise and in relative terms, recipients received a decrease in wealth of 33% (loss/expected amount: $\$50/\$150 = 33\%$). In the honest-absolute condition, recipients received 50% more than they would have received if their partners had been dishonest. In the honest-relative condition, recipients would have received \$75 if Pat had been dishonest, so received an increase in wealth of 33% (gain/expected amount: $\$25/\$75 = 33\%$). Moreover, to remove social comparisons, the scenario included information only about recipients' earnings but not their partners' earnings.

Recipients were then given information that they could hypothetically spend money to reward Pat (in the honesty condition) or punish Pat (in the dishonesty condition) at a low or at a high monetary cost. These two levels of cost were chosen based on the past design by Brandts & Charness (2003). The instructions for rewarding at a low cost (*italicized and not parenthesized*) and at a high cost (*italicized and parenthesized*) are presented below:

You have a one-time opportunity to reward Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to giving money to Pat at a *1:10 (1:1)* ratio. In other words, for every *10 cents (1 dollar)* you spend, you reward Pat 1 dollar. You can reward up to \$100. After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to add money to your bank.

Recipients were able to hypothetically give money in the honest condition to their partner or subtract money in the dishonest condition from their partner. In doing so, the participants faced a cost to react equal to a tenth of the impact on Pat or equal to the total impact on Pat.

Dependent measures

Following these instructions, participants were asked how they wanted to respond to their partner. The possible punishment/reward amounts were presented in an 11 point Likert-type scale and ranged from \$0 to \$100, in increments of \$10. The costs to respond were also included in the scale. For example, in the dishonest-high cost condition, a participant who chose to punish \$10 checked the option 'Punish \$10 (at a cost of \$10)'. A participant who chose to punish \$0 checked the option 'Do not punish'.

Choice to respond (frequency). The choice to respond or not to respond to their partner was coded as a dichotomous variable, with a response defined as the participant punishing or rewarding any amount and no response defined as the participant not to punishing or rewarding.

Amount of response (intensity). The monetary amount participants rewarded or punished their partners was measured. Two different analyses were performed, one including and the other excluding zero responses.

Results

The main theoretical objective was to determine if recipients punished deception more than they rewarded honesty when the acts were equivalent in absolute and relative terms. The responses were compared in terms of frequency and intensity.

Gender covariate

Past data has suggested gender differences in trusting and reciprocal behaviors, with women reciprocating more than men (Chaudhuri & Gangadharan, 2007; Croson & Buchan, 1999; Snijders & Keren, 2001). As a result, I tested for gender differences in the Experiment 1 analyses. There were no significant interactions with gender in the following analyses (F 's < 1). However in line with past research, women responded more frequently ($\chi^2 = 4.60, p < .05$) and intensely (including zero responses, $\chi^2 = 3.20, p < .05$) than men. Moreover, because the gender breakdown differed by condition (see Table 1), I conducted two sets of analyses, one with and one without gender as a covariate, and obtained the same pattern of results. The reported analyses for Experiment 1 did not include gender as a covariate.

Choice to respond (frequency)

The percentage responding and frequency counts are displayed by condition in Figure 2. I conducted a binary logistic regression to test the effect of partner's behavior and cost on the choice to respond. The explanatory terms included partner's behavior (categorical: deception, honesty-absolute, honesty-relative), cost to respond (dichotomous: low, high), the interaction between the two.

The behavior of partner x cost interaction was not significant, $\chi^2(2, N = 107) = .07, p = ns$. There was a marginally significant effect for cost, $\chi^2(1, N = 107) = 2.66, p = .10, \text{Exp}(B) = 4.43$. Recipients were more likely to respond when costs were low (77%) than when costs were high (57%). More importantly, there was a significant effect for partner's behavior, $\chi^2(2, N = 107) = 9.33, p < .01$. Each act of honesty was compared to the act of deception, the first comparing deception and honesty equal in terms of absolute magnitude and the second comparing deception and honesty equal in terms of relative magnitude.

Comparing by absolute standards, recipients of honesty rewarded more often (86%) than recipients of deception punished (40%), $\chi^2(1, N = 107) = 9.01, p < .01, \text{Exp}(B) = 11.25$.

Comparing by relative standards, recipients of honesty rewarded more often (74%) than recipients of deception punished (40%), $\chi^2(1, N = 107) = 4.32, p < .05, \text{Exp}(B) = 5.00$. While this was not the central theoretical question, the honesty-absolute and honesty-relative conditions were also compared and no significant difference emerged, $\chi^2(1, N = 107) = 1.13, p = .28, \text{Exp}(B) = .44$.

Amount of response, including zero responses (intensity)

Intensity of response was analyzed in a 3 (behavior of partner: deception, honesty-absolute, honesty-relative) x 2 (cost to respond: low, high) between participants analysis of variance (ANOVA). The interaction was not significant ($F < 1$). There was a main effect for cost to respond, with low cost participants ($M = 57.41, SD = 40.31$) responding more intensely to their partner's behavior than high cost participants ($M = 21.57, SD = 23.52$), $F(1, 101) = 35.01, p < .001$.

There was also a main effect for the partner's behavior, $F(1, 101) = 5.03, p < .01$. The planned contrasts suggested that, given equivalent absolute levels of deception and honesty, recipients rewarded ($M = 47.57, SD = 29.57$) more intensely – almost twice as much – than they punished ($M = 26.57, SD = 38.73$), $F(1, 101) = 9.01, p < .01$. Given equivalent relative levels of deception and honesty, recipients also rewarded ($M = 45.57, SD = 42.16$) more intensely – almost 1.5 times as much – than they punished, $F(1, 101) = 5.81, p = .01$. The two honesty conditions did not differ significantly in levels of reward, $F(1, 101) < 1, p = \text{ns}$. The means and number of participants are displayed by condition in Figure 3.

Amount of response, excluding zero responses (intensity)

I also compared the intensity of responses of the participants who chose to respond. As a result, 32 participants who did not respond were excluded, leaving 72 participants in the analysis. Another 3 (behavior of business partner: deception, honesty-absolute, honesty-relative) x 2 (cost to respond: low, high) between participants ANOVA was performed (see Figure 4).

The interaction and the main effect for partner's behavior were not significant ($F < 1$). Cost to respond was the only significant effect, with low cost participants ($M = 74.77$, $SD = 28.26$) responding significantly more to their partner's behavior than high cost participants ($M = 36.90$, $SD = 19.48$), $F(1, 66) = 35.40$, $p < .001$. The small number of participants left in the deception condition following the exclusion of zero responses may have led to the non-significant effect for partner's behavior (see Figure 4). Specifically, over 57% of deceived recipients chose not to respond, leaving only 4 participants in the high cost-deception condition and 10 participants in the low cost-deception condition. This suggests that additional data are needed before making strong conclusions.

Discussion

Experiment 1 presented a scenario in which recipients of honesty and deception could monetarily reward and punish the actor and provided support for the central theoretical argument that recipients are likely to reward honesty more than they punish deception. By removing the asymmetries of past experimental designs and by providing equivalent levels of deception and honesty in both absolute and relative terms, Experiment 1 provided a clearer comparison of how recipients might punish versus reward. Contrary to the overwhelming consensus within the economics literature (Abbink et al., 2000; Brandts & Charness, 2003; cf. Fehr & Gächter, 2000b;

Offerman, 2002) and the assumptions with the psychology literature (Baumeister et al., 2001; Rozin & Royzman, 2001), the results suggest that as a whole, participants would reward more frequently and intensely than they punish.

Within the scenario, recipients responded less frequently and intensely when costs were high, regardless of the Pat's behavior. While Experiment 1 provides an initial comparison of how differing levels of cost influence the distribution of punishments and rewards, additional tests might investigate how multiple levels of costs might differentially influence punishment and reward behavior. Moreover, recipients responded within a scenario that included anticipated responses, which may differ from responses that involve actual deception and honesty and monetary repercussions. Experiment 2 used a different methodology to compare punishments and rewards with real monetary stakes.

CHAPTER 2

Experiment 2: Why recipients reward more than they punish

Experiment 2 attempted to replicate Experiment 1. To increase the validity of the results, it also involved behavioral responses with actual monetary incentives. Experiment 2 extends the findings of Experiment 1 in three additional ways. First, Experiment 2 added a no cost condition to determine whether the effects of cost may be more nuanced than the two levels established in Experiment 1. Second, Experiment 2 included three psychological mediators, with felt obligation, an affective measure (e.g., negative affect), and a cognitive measure (e.g., perception of trust) to shed insight into how and why recipients punish deception and reward honesty. Past theorizing suggests that felt obligation will encourage positive reciprocity more than it encourages negative reciprocity (Gouldner, 1960; Kunz & Woolcott, 1976; Pillutla et al., 2003). Thus, I tested whether recipients of honesty felt more obligated to reward than recipients of deception felt obligated to punish. I also tested how negative emotions (i.e., affect) and perceptions of trust (i.e., cognition) influenced both felt obligation and decisions to punish and reward. Finally, Experiment 2 tested how these potential mediators impacted responses in a subsequent, unrelated context. The supporting literatures for these extensions are discussed in the following sections.

Affective- and cognitive-based responses

Affect and cognition have long been recognized by moral philosophers (see Haidt, 2001; Hauser, 2006, for reviews) and social psychologists (see Cacioppo & Gardner, 1999; Cacioppo, Gardner, & Berntson, 1999, for reviews) as precursors to judgment and decision-making. Affect and cognition have been delineated by their evaluative nature, with affect tied to valenced

evaluative processing and cognition tied to perceptual processing. Affective responses have been discussed in terms of emotions (e.g., I am happy) and preferences (e.g., I like brick-oven pizza). In contrast, cognition has been described in terms of information processing, being more involved in acts of deliberation, identification and discrimination (e.g., determining whether a pizza is brick-oven or deep dish) (Cacioppo & Gardner, 1999; Cacioppo et al., 1999; Zajonc, 1980, 1984).

A steady discussion has emerged in the past three decades, with debate on whether affect can influence judgment and behavior without the presence of cognition. Classic theories in social psychology assumed cognitions were necessary in judgments and behavior and that affect could not arise without prior cognitions (Lazarus, 1982). More recent perspectives suggest that affective and cognitive processes are housed in separate and relatively independent systems that often function in conjunction, but also can influence decisions separately (Cacioppo & Gardner, 1999; Cacioppo et al., 1999; Zajonc, 1980, 1984). Neurophysiological discoveries support this premise, finding that affective responses are housed in one area in the brain and cognitive appraisals are housed in another area (see Zajonc, 1984, for a review; Zola-Morgan, Squire, Alvarez-Royo, & Clower, 1991). Thus, affect and cognition may differentially impact decisions to positively and negatively reciprocate.

Punishment and Emotions

Recent evidence suggests that punishment decisions are driven by negative emotions. Loewenstein (1996) suggests that certain emotions interfere with the rational goal of monetary maximization and that emotions, such as anger, seem to play a part in the decision to punish even at a self-cost. Researchers report that recipients punished more when angry or reporting negative

emotions (Bosman, Sutter, & van Winden, 2005; Brosig et al., 2003; Pillutla & Murnighan, 1996). For example, Pillutla and Murnighan (1996) cite spite and anger at receiving small ultimatum offers as the rationale for high rejection rates. Brosig and her colleagues (2003) recently compared ‘hot’ games, in which individuals react automatically to another individual’s behavior, to ‘cold’ games, in which individuals pre-program a strategy prior to interaction. Recipients with low costs were more willing to punish during hot games than during cold games. The authors suggest that negative emotions increase punishment.

In the new field of neuroeconomics, researchers have that found brain activity in the anterior insula, the area associated with negative emotions, and the dorsolateral prefrontal cortex, the area associated with cognition, are heightened when responders received an unfair ultimatum offer (e.g., less than \$5-\$5 split). Moreover the anterior insula was significantly heightened, but not the dorsolateral prefrontal cortex, when responders *rejected* unfair offers, suggesting that emotions play a larger role than cognition in punishment behavior (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003).

Cognitive appraisal and reward

In contrast, there is evidence that reward behavior results more from cognitive appraisals than from emotions. Two forms of interpersonal trust have been defined: cognitive- and affective-based trust (McAllister, 1995). Cognitive-based trust is grounded more in evidence-based perceptions, such as indications of trustworthy behavior, whereas affect-based trust arises from positive evaluations of others (McAllister, 1995). Cognitive-based trust has been tied to reward behavior, where people reward those that have acted in a trustworthy and cooperative manner (McCabe et al., 1998; McCabe, Rigdon, & Smith, 2003; Pillutla et al., 2003). Evidence

suggests that recipients feel obligated to reward trustworthy actors. More specifically, Pillutla and his colleagues (2003) found that trusting behavior by actors led recipients to feel obligated to reciprocate and in turn resulted in recipients rewarding actors.

Neuroeconomics data also supports that perception of trust influences reward behavior. The dorsolateral prefrontal cortex, the region recognized for deliberation & cognition, is heightened when recipients accept ultimatum offers (Sanfey et al., 2003), which results in both parties receiving the monetary amount proposed by the actor. This area is heightened when intending to reward in trust games (King-Casas et al., 2005), suggesting that the perceived trustworthiness of the actor – a cognitive factor – will determine reward amounts.

How emotions and cognitions drive indirect responses

Considerable research on responses to deception and honesty has focused either on direct reciprocation of negative and positive behavior (Fehr & Gächter, 2000b) or on third-party punishment, when wrongdoers are punished by those not harmed by the wrongdoer (Kahneman, Knetsch, & Thaler, 1986; Sober & Wilson, 1998; Fehr & Fischbacher, 2004).

Recent explorations give insight to how recipients of deception and honesty might act in a subsequent, unrelated context, with incidental emotion from one situation influencing normatively unrelated decisions (Forgas, 1995; Loewenstein, 1996; N. Schwarz, 1990). The response when the four officers accused of beating Rodney King were acquitted vividly illustrate how anger from experiencing injustice might generate responses in a subsequent, unrelated context. The 1992 verdict sparked outrage from the community and riots in Los Angeles leaving 50 people dead, 2000 injured and 800 buildings burned (Cannon, 1999).

Consistent with this notion, empirical evidence suggests that people have a greater willingness to engage in deviant behavior when people feel an outcome is unjust. For example, perceptions that a particular law is unjust can lead to greater self-reported willingness to engage in lawbreaking (e.g., petty theft) in a subsequent, unrelated context (Nadler, 2005, see also Robinson & Darley, 1995).

Moreover, the negative emotions resulting from feelings of injustice are thought to mediate this process. For example, Mullen & Nadler (2007) found evidence for a moral spillover effect, such that people are more willing to engage in deviant behavior when outcomes oppose rather than support their moral standards. Results revealed that relative to just outcomes, people who thought that outcomes were unjust were angrier and were more likely to keep a borrowed pen. Since affect can spill over to other contexts, it seems particularly likely that the negative emotions felt after deception might result in deviant behavior – theft of a pen – in an unrelated context.

There is less concrete evidence that cognitions will spill over to unrelated situations. However, the stereotyping literature provides some indication that people might act in a more trustworthy manner after being in contact with people they perceive as trustworthy. Contact with and even mere exposure to a stereotyped group member increases behaviors that are stereotypical of that group (Dijksterhuis, Aarts, Bargh, & van Knippenberg, 2000; Wheeler & Petty, 2001). For example, the more time an individual spends with the elderly, the more forgetful that individual becomes after being primed with the elderly stereotype (Cesario, Plaks, & Higgins, in press). Thus, recipients who view those rewarding them as trustworthy may act in a more trustworthy manner themselves (i.e., less deviant behavior like theft of pens).

Overview

Experiment 2 included behavioral measures of punishment and reward and post-behavior psychological measures to help explain how and why recipients reward and punish. Figure 1 outlines the model proposed for Experiment 2. I suggested that affective responses (e.g., negative emotions) mediated the relationship between deception and punishment, whereas cognitive appraisals (e.g., perceptions of trust) mediated the relationship between honesty and reward. Moreover, I suggested that felt obligation explained why recipients rewarded honesty more than they punished deception in Experiment 1. I posited that perceptions of trust lead to felt obligation, which would help explain why recipients rewarded honesty more than they punished deception. Finally, Experiment 2 tested how these mediators also impacted responses in a subsequent, unrelated context. I suggested that negative emotions increased and greater perceptions of trustworthiness decreased deviant behavior (i.e., pen theft).

Method

Experiment 2 was a controlled laboratory experiment, with a similar design to Experiment 1, in which I manipulated the predictor variables (e.g., behavior of partner; cost to respond) and measured punishment or reward behavior, psychological responses, and subsequent behavior.

Participants

Participants were 184 undergraduate students (83 men and 101 women) from a major Midwestern University. The students were recruited via email, paid a \$5 show up fee, and were told they might earn additional money during the experiment. The experiment lasted approximately 30 minutes.

Design and Procedures

The experiment had a 3 (behavior of other player: deception, honesty-absolute, honesty-relative) x 3 (cost to react: no, low, high) between-subjects design. When participants arrived for the experiment, they were taken to separate breakout rooms and filled out consent forms. They were informed that the experiment consisted of two stages and that they would be interacting with the same person in both stages. Participants were told that the interactions consisted of a series of decisions and that they would never meet or know the identity of the other participant. Participants were told that in addition to the \$5 participation fee, which they received in all cases, they would also receive the actual amount of money that they accumulated in the experiment

Stage 1. In Stage 1, participants played a modified version of the deception game (Gneezy, 2005). Each participant ostensibly played a game with another player who had the ability to deceive the participant. Responses were pre-programmed so that the participant believed that he or she was the recipient of either honesty or deception, depending on the condition. The full set of instructions given to the participant for Stage 1 can be found in Appendix 2.

In Stage 1, the participant chose between two options: Option A and Option B. Each option resulted in a monetary payment for the participant and the other player. The participant did not have information about the amounts in each option, but did know that one of the options would earn him/her more than the other option (e.g., Option X earns the participant more than Option Y), whereas the reverse options would earn the other player more (e.g., Option Y earns the other player more than Option X). The participant was also told that the other player knew the actual amounts in each option and would

send him/her one of two messages, “Option A earns you more than Option B” or “Option B earns you more than Option A”, one of which was misleading and the other of which was accurate.

After reading the instructions, the participant received the message from the other player. The pre-programmed message was “Option A will earn your more money than Option B.” Participants were then given the choice between Option A and Option B. After making a choice, the participant was given a message pre-programmed according to condition.

In the *deception (honest)* condition, participants that believed Player A (i.e., chose Option A) received the following message:

Now that you have made your decision, we can reveal that the true message was: *Option B (Option A) will earn you more than Option A (Option B). Because you chose Option A, you received \$4 in Stage 1. You would have received 50% more (less) if you had chosen Option B.*

An additional set of messages were created when participants chose not to believe Player A (i.e., chose Option B). In the *deception (honest)* condition, participants who did not believe Player A received the following message:

Now that you have made your decision, we can reveal that the true message was: *Option B (Option A) will earn you more than Option A (Option B). Because you chose Option B, you received \$4 in Stage 1. You would have received 50% less (more) if you had chosen Option A.*²

Similar to Experiment 1, participants received the same amount across all conditions and did not know the actual allocation the other players received.

Stage 2. Stage 2 was a modified dictator game (for overviews of dictator games, see Camerer, 2003; Forsythe, Horowitz, Savin, & Sefton, 1994) in which participants were able to add or subtract money from their counterparts' endowments. Participants were told that they would interact with the same person with whom they interacted in Stage 1 and were reminded that Stage 2 was the final stage of the experiment. Moreover, participants were told that their counterparts would not have the option to subtract from or add to their endowment. Thus, participants were informed that they would be the only ones making decisions in Stage 2.

The participants were told that the other player had a positive amount in his/her bank. In the cost conditions, participants were able to use the \$4 earned in Stage 1 to add or subtract money from their counterpart player's endowment if they wished. Participants faced a cost to react equal to zero, a tenth of the impact on the counterpart, or the total impact on the counterpart. Thus, depending on the condition, adding or subtracting \$4 from the counterpart's endowment cost the player nothing, \$0.40, or \$4. The decision sheet with instructions can be found in Appendix 3. An example of the instructions in the reward-low cost condition is as follows:

You have a one-time opportunity to reward Player A, your choice will require you to spend your own money from Stage 1. For every 10 cents you spend, you can reward Player A 1 dollar. Rewards can be at most 4 dollars.

After recording their decisions, participants completed post-experimental questionnaires to assess their perceptions of their counterparts, the process, their emotions and moods, and various demographic variables. At this point, participants were paid (total amount earned = \$9 – amount spent to respond).

Dependent measures

Punishment or reward behavior. Like Experiment 1, the two dependent measures were (a) *frequency*: the choice to respond or not to respond to the other player's behavior and (b) *intensity*: the amount participants rewarded or punished the other players. Intensity was analyzed including and excluding zero responses.

Psychological measures. After their behavioral choices, participants responded to three psychological measures: a measure of cognitive response, perceptions of trustworthiness; a measure of emotional response, level of negative affect; and felt obligation. The three measures were adapted from measures developed by Pillutla et al. (2003). Please see Appendix 4 for the post-questionnaire containing the psychological items.

Perceptions of trustworthiness were measured by four items. A score was created by averaging the responses. Ratings and scores ranged from 1 to 7 (e.g., "I consider the other player to be extremely trustworthy/untrustworthy", 1 = "completely untrustworthy" to 7 = "completely trustworthy"). The alpha coefficient was .88.

Negative affect was measured by six items. A score was created by averaging the responses. Ratings and scores ranged from 1 to 7 (e.g., "After you found out the information about the other player's choice did you feel", 1 = "completely happy" to 7 = "completely unhappy"). The alpha coefficient was .82.³

Finally feelings of obligation were measured using a single-item scale. The measure ranged from 1 to 7 (e.g., "When making the decision in Stage 2, I felt", 1 = "not at all obligated to react" to 7 = "completely obligated to react").

Subsequent behavior in an unrelated context. A cup filled with 7 high-quality roller ball pens was placed on the table for use in filling out the final questionnaire. Participants were told that once they had finished the questionnaire, they could leave at their leisure. After the participant left, the researcher counted the number of pens in the cup to determine whether the participant took a pen home with them.

Results

Behavior of other player manipulation check

After Stage 1, I asked participants to identify whether the other player had been deceptive or honest to ensure that the manipulation of the other player's behavior worked. Eight participants failed the manipulation check and were dropped from further analysis, leaving 176 (79 men and 97 women).

Gender covariate

Similar to Experiment 1, I tested for gender differences in the Experiment 2 analyses. There were no significant interactions with gender, $F < 1$. Also similar to Experiment 1, the gender breakdown differed within certain conditions (see Table 2). Thus, I conducted two sets of analyses, one with and one without gender as a covariate. I included the gender covariate in two analyses, the frequency analysis and the negative affect and trust moderated-mediation analyses for intensity, as the results were improved by its inclusion. The other analyses reported did not include gender as a covariate.

Choice to respond (frequency)

The choice to respond was analyzed in a binomial logistic regression. The independent variables included the other player's behavior (categorical: deception, honesty-absolute, honesty-

relative), cost to respond (continuous: no, low, high), the interaction between the two, and gender. Cost to respond was used as a continuous variable (0, .10, or \$1 to punish/reward \$1) rather than categorical to provide a more precise measure. The interaction was significant, $\chi^2(2, N = 176) = 6.05, p < .05$. The percentage responding and frequency counts are displayed by condition in Figure 5.

Similar to Experiment 1, there was an effect for the behavior of the other player, $\chi^2(2, N = 176) = 13.88, p = .001$. Recipients of honesty rewarded more often in both the absolute (87%), $\chi^2(1, N = 176) = 9.27, p < .01, \text{Exp}(B) = 5.52$, and relative conditions (92%), $\chi^2(1, N = 176) = 23.01, p < .01, \text{Exp}(B) = 23.01$, than recipients of deception punished (35%). The two honesty conditions did not differ, $\chi^2(1, N = 176) = 1.39, p = .24$.

The interaction was decomposed to understand the effects of cost by the behavior of the other player. The frequency of punishment was significantly affected by increasing costs. The odds for punishment decreased significantly as cost increased, $\chi^2(1, N = 176) = 11.28, p < .001, \text{Exp}(B) = .09$. Planned chi-squared tests revealed that both no cost (45%), $\chi^2(1, N = 176) = 5.68, p < .05$, and low cost (60%), $\chi^2(1, N = 176) = 13.91, p < .001$, recipients of deception punished more than high cost recipients of deception (10%). No cost and low cost recipients of deception did not significantly differ in punishment frequencies $\chi^2(1, N = 176) < 1$. Thus, punishments occur approximately half the time when costs were low or non-existent, but they rarely occurred when costs were high. In contrast, reward decisions were essentially unaffected by costs in the honesty-absolute condition (no cost, 93%; low cost, 77%; high cost, 90%), $\chi^2(1, N = 176) < 1$, and the honesty-relative condition (no cost, 100%; low cost, 93%; high cost, 83%), $\chi^2(1, N =$

176) < 2, $p = .18$.⁴ Moreover, the results suggested that rewards are frequent, with over three-fourths of recipients rewarding in all of the honesty conditions.

Amount of response, including zero responses (intensity)

Intensity was analyzed in a 3 (behavior of partner: deception, honesty-absolute, honesty-relative) x 3 (cost to respond: no, low, high) between participants ANOVA (see Figure 6).

Although the behavior x cost interaction was not significant, $F(4, 167) < 1$, the main effect for partner's behavior was significant, $F(1, 167) = 15.62, p < .001$. Planned contrasts revealed that participants rewarded honesty more than they punished deception ($M = .92, SD = 1.46$), in both absolute ($M = 2.24, SD = 1.44$), $F(1, 167) = 25.22, p < .001$, and relative terms ($M = 2.24, SD = 1.22$), $F(1, 167) = 18.83, p < .001$. Participants in the honesty-absolute and honesty-relative conditions did not significantly differ in reward amounts, $F < 1$.

There was also a main effect for cost, $F(1, 167) = 8.04, p < .001$. There was no significant difference between the no cost ($M = 2.10, SD = 1.61$) and low cost participants ($M = 2.11, SD = 1.57$), $F < 1$, but high costs led to significantly less punishments and rewards ($M = 1.00, SD = 1.15$) than no costs, $F(1, 167) = 10.32, p < .01$, and low costs, $F(1, 167) = 12.72, p < .001$.

Amount of response, excluding zero responses (intensity)

A 3 (behavior of business partner: deception, honesty-absolute, honesty-relative) x 2 (cost to respond: no, low, high) ANOVA was performed with participants who did not punish or reward excluded (see Figure 7 for means and frequencies per condition). Fifty-seven participants did not reward or punish; they were excluded, leaving 119 participants in the analysis. The interaction and main effects were not significant ($F < 1$), suggesting that recipients who punished

and rewarded did so in a similar manner, regardless of cost. However, similar to Experiment 1, this analysis included few participants in the deception condition and should be interpreted with caution until more data is collected. Next, I examined the psychological measures underlying punishment and reward behavior. Table 3 provides means and correlations for the entire sample on all of the variables reported in Experiment 2.

Level of Trust

I predicted that recipients would trust actors who performed honest acts more than deceptive acts. Recipients' trust was analyzed in a 3 (behavior of the other player: deception, honesty-absolute, honesty-relative) x 3 (cost to respond: no, low, high) between participants ANOVA. There was a main effect for the behavior of the partner, $F(2, 167) = 88.53, p < .001$. As expected, participants trusted honest more than dishonest actors ($M = 3.38, SD = 1.01$) in both absolute ($M = 5.25, SD = .89$), $F(1, 167) = 137.53, p < .001$, and relative terms ($M = 5.37, SD = .87$), $F(1, 167) = 113.18, p < .001$. There was also a main effect for cost, $F(1, 167) = 4.06, p < .05$. No cost participants ($M = 4.82, SD = 1.19$) trusted actors more than low cost recipients ($M = 4.38, SD = 1.26$), $F(1, 167) = 6.96, p < .01$.

Negative affect

I predicted that recipients would be unhappier when actors were dishonest than when they were honest. Higher numbers reflected more negative affect. Participants' negative affect was analyzed in a 3 (behavior of the other player: deception, honesty-absolute, honesty-relative) x 3 (cost to respond: no, low, high) between participants ANOVA. There was a main effect for the behavior of the other player, $F(2, 167) = 88.53, p < .001$. As expected, participants had more negative affect following honesty than dishonesty ($M = 3.38, SD = 1.01$) in both absolute ($M =$

5.25, $SD = .89$), $F(1, 167) = 137.53, p < .001$, and relative terms ($M = 5.37, SD = .87$), $F(1, 167) = 113.18, p < .001$.

Felt obligation to respond

I predicted that participants would feel more obligated to reciprocate positive acts than negative acts. The measure was reverse coded with higher numbers reflecting greater felt obligation to respond. Participants' felt obligation was analyzed in a 3 (behavior of the other player: deception, honesty-absolute, honesty-relative) x 3 (cost to respond: no, low, high) between participants ANOVA. There was one significant effect, a main effect for the behavior of the other player, $F(2, 167) = 4.05, p < .05$. As expected, participants felt more obligated to reward honesty (absolute: $M = 3.63, SD = 1.58, F(1, 167) = 5.55, p < .05$; relative: $M = 3.70, SD = 1.60, F(1, 167) = 5.95, p < .05$) than to punish deception ($M = 2.86, SD = 1.91$).

Trust and negative affect moderated-mediation analyses (intensity)

Because there were no significant differences between responses in terms of frequency, intensity, and the psychological measures in the two honesty conditions, all of the subsequent analyses collapsed these two conditions. To test the effects of negative affect and trust on the amount punished and rewarded, two separate moderated mediation analyses were run using the procedures outlined by Preacher, Rucker, and Hayes (2007) and Shrout and Bolger (2002), one with trust and the other with negative affect as the mediator. A moderated mediation results when a treatment effect is mediated differently as a function of some moderator variable. I conducted a type of moderated mediation in which the independent variable also serves as the moderator. This type of moderated mediation occurs when the effect of the mediator (M) on the

dependent variable (Y) is altered by the independent variable (X), i.e., the mediator and the independent variable interact to cause the outcome (Muller, Judd, & Yzerbyt, 2005).

There are two steps in this type of moderated mediation; the first is to establish a relationship between the independent variable and the mediator and the second is to establish that the relationship between the mediator and dependent variable is moderated by the independent variable. This is statistically tested by the impact of the interaction between the independent variable and mediator on the dependent variable.⁵

In the first moderated mediation, I predicted that trust mediated the relationship between honesty and reward, but did not mediate the relationship between deception and punishment. In other words, the actors' behavior (X) should predict how much recipients' trusted the actors (M), with recipients (naturally) trusting honest actors more than they did dishonest actors. In turn, trust levels (X) should be positively associated with intensity of responses (Y) *only* when actors were honest and not when they were dishonest. Therefore, the relationship between trust levels (M) and intensity of responses (Y) should depend on the actors' behavior (X): greater levels of trust should be positively associated with reward amounts, but decreased trust should not be associated with punishment amounts (see Figure 8a).

The second moderated mediation used the negative affect measure as the moderator/mediator (M). I predicted that negative affect would mediate the relationship between deception and punishment, but not the relationship between honesty and reward (see Figure 8b). Table 4 outlines the regression estimates for both moderated mediation analyses.

Trust as a mediator. As the first step of the trust moderated mediation, I examined the relationship between the other player's behavior and trust with an ordinary least squares (OLS)

regression, controlling for cost and gender. Regressing the behavior of the player (deceptive = 0, honest = 1) on trust was significant, $\beta = .71, p < .001$: honesty, as compared dishonesty, led to perceptions of the other player as more trustworthy.

Next, I investigated if the relationship between trust and intensity was moderated by the other player's behavior. To test for moderation, I conducted an OLS regression that included terms for trust perceptions, the other player's behavior, the interaction between trust and the other player's behavior (reflecting a moderation effect), and gender. I also controlled for the effect of negative affect and the negative affect x other player's behavior interaction. Controlling for negative affect removed the explanatory variance of affect and allowed for a test of the independent impact of the cognitive variable (e.g., trust perceptions) on intensity. The interaction was marginally significant, $\beta = .20, p < .10$, suggesting that the association between trust perceptions and intensity depended on whether the actor had been honest or deceptive to the recipient.

To test the significance of the indirect effects, I followed the procedure described by Holmbeck (2002). I ran two OLS regressions: one generating the simple slope for the deception sample and one generating the simple slope for the honesty sample. Results suggested that the more that the honest players were perceived as trustworthy, the more recipients rewarded, $\beta = .27, p = .05$. However, viewing a deceptive partner as more untrustworthy was not positively associated with the amount of punishment, $\beta = -.12, p = ns$ (see Figure 9).

Negative affect as a mediator. A similar analysis was conducted to test the mediating and moderating impact of negative affect. I hypothesized that the actors' behavior predicted recipients' negative affect: recipients were (naturally) unhappier when actors were dishonest than

when they were honest. In turn, more negative affect should be positively associated with the amount of punishment, but less negative affect should not be associated with the amount of reward.

The first step of the moderated mediation regressed partner's behavior (deception = 0, honesty = 1) on negative affect, controlling for cost, and gender. This regression was significant, $\beta = -.48, p < .001$, showing that dishonesty, as compared to honesty, generated greater negative affect.

The next step was to test the negative affect x behavior interaction, controlling for cost, gender, and the cognitive variables (e.g., trust, trust x behavior interaction). The interaction was significant, $\beta = -.21, p < .05$. The interaction suggested that negative affect was positively associated with the punishment of deceptive actors, marginally, $\beta = .26, p = .10$. However, recipients who were happier about the actors' honesty did not bestow greater rewards, $\beta = -.12, p = ns$ (see Figure 10).

Felt obligation moderated-mediation analyses

To shed light on why participants reward more than they punish, two additional moderated mediation analyses were run using procedures from James and Brett (1984). The main objective was to determine the psychological mechanism driving feelings of obligation to respond. I predicted that felt obligation mediated the trust-reward relationship, but not the negative affect-punishment relationship.

In the first moderated mediation analysis, I tested whether felt obligation mediated the relationship between trust and intensity after honesty but not after deception. To test this idea, I first looked at the pattern of correlations separately after deception and honesty. Trust was

significantly correlated with amount of reward, $r(107) = .23, p < .05$, but not with amount of punishment, $r(69) = -.10, p = .41$. As a result, a mediation analysis was performed looking at the effect of felt obligation on the relationship between trust and reward behavior, controlling for cost and negative affect.

Three regressions were performed in the honesty condition, (1) regressing trust on intensity of reward, (2) regressing trust on felt obligation, (3) and regressing both trust and felt obligation on amount of reward. Regressing trust on intensity of reward was significant, $\beta = .39, p < .01$, suggesting that trust was positively associated with intensity of reward. Regressing trust on felt obligation was significant, $\beta = .51, p < .001$, suggesting that higher felt obligation was positively associated with greater rewards. Finally, the third regression tested for mediation: when controlling for obligation, $\beta = .34, p < .001$, the relationship between trust and reward was no longer significant, $\beta = .08, p = .42$. The Sobel test was significant, $z = 2.86, p < .01$ (see Figure 11).

In the second moderated mediation analysis, I tested to see whether felt obligation mediated the relationship between negative affect and intensity after honesty and deception. I looked at the pattern of correlations separately after deception and honesty. Negative affect was not significantly correlated with punishment, $r(69) = .13, p = .28$, but was correlated with amount of reward, $r(107) = .29, p < .01$. As a result, the first step of a mediation analysis looked at the effect of felt obligation on the relationship between negative affect and reward behavior, controlling for cost and trust. However, the regression of negative affect on amount of reward was not significant, $\beta = -.15, p = ns$. As a whole, results support that trust mediated the honesty-

reward relationship, negative affect mediated the deception-punishment relationship, and obligation mediated the trust-reward relationship.

Subsequent behavior (pen theft)

I tested how the other player's behavior and cost influenced subsequent behavior, namely pen theft. Regressing the other player's behavior, cost, and the behavior x cost interaction on pen theft revealed only one marginally significant main effect for cost, $\chi^2(1, N = 176) = 2.76$, $\text{Exp}(B) = .34$, $p = .10$. Results suggested that as cost increased, recipients were less likely to take a borrowed pen (see Figure 12 for frequencies).

Negative affect and trust moderated-mediation analyses (pen theft)

Next, I performed two additional moderated mediation analyses to see if trust and negative affect influenced subsequent behavior, again using the procedures outlined by Preacher, Rucker, and Hayes (2007) and Shrout and Bolger (2002). Subsequent behavior was operationalized as whether recipients took a pen with them after the completion of the experiment. Table 5 outlines the regression estimates of both subsequent behaviors in moderated mediation analyses.

Trust as a mediator. The first moderated mediation included trust as the mediator. The first step, establishing the relationship between partner's behavior and trust was completed in the previous analyses. To test if the effects of trust were moderated by partner's behavior, I then conducted a logistic regression that included terms for the main effects for trust, the other player's behavior, along with the interaction between the two on the odds ratio of subsequent behavior. I also controlled for cost to respond, negative affect, and the negative affect x behavior interaction.

There was a significant interaction, $\chi^2(1, N = 176) = 4.09, p < .05$. To probe the interaction, I ran two separate logistic regressions to understand the effects of the other player's behavior as a moderator of subsequent behavior. Following honesty, each additional unit of trust decreased the odds of stealing a pen, $\chi^2(1, N = 176) = 7.03, p < .01, \text{Exp}(B) = .43$. Following deception, trust did not influence the odds of stealing a pen $\chi^2(1, N = 176) < 1$. These findings are depicted in Figure 13a.

Negative affect as a mediator. I performed a separate moderated mediation analysis with negative affect as the mediator. The influence of partner's behavior on negative affect was previously established. To test if the effects of negative affect were moderated by partner's behavior, I conducted a logistic regression analysis that included terms for the main effects for negative affect, partner's behavior, and the interaction between the two on the odds ratio of subsequent behavior. I also controlled for cost to respond, trust, and the trust x behavior interaction. The interaction was marginally significant $\chi^2(1, N = 176) = 3.27, p = .07$. The results suggested that negative affect increased theft following deception, however, the results were not significant, $\chi^2(1, N = 176) = 2.03, p = .15, \text{Exp}(B) = 2.04$. The trends suggested that more negative affect increased theft following deception. Following honesty, however, happiness did not influence the odds of stealing a pen, $\chi^2(1, N = 176) = .13, p = \text{ns}, \text{Exp}(B) = 1.13$. These findings are depicted in Figure 13b.

Suppressor effects. Whereas no direct link was found between the other player's behavior and pen theft, in the case of moderated mediation, it is possible that the complexity of interactions between key mediators and moderator (the other player's behavior) suppressed the influence of the independent variable (the other player's behavior) on the dependent variable

(pen theft) (Shrout and Bolger, 2002). Thus, the statistical relationship between the other player's behavior and pen theft may be stronger with the inclusion of the mediators. Consistent with this conjecture, after controlling for the suppressing mediating variables and covariates, an association between partner's behavior and theft emerged, with honest recipients marginally more likely to steal a pen than dishonest recipients, $\chi^2(1, N = 176) = 3.92, p = .06, \text{Exp}(B) = 3.19$. This result, still tentative and without strong theoretical grounding, should be studied more in the future.

Discussion

Experiment 2 was a behavioral replication of Experiment 1 with actual monetary stakes. Results replicated Experiment 1's results, confirming that recipients rewarded honesty more frequently and intensely than they punished deception. Recipients rewarded more than punished whether the amount of deception and honesty were equivalent in absolute or in relative terms.

Experiment 2's results also suggested that costs might impact the frequency of punishments and rewards in different ways. For instance, recipients continued to reward honesty even as costs increased. They punished less frequently, however, as costs increased. Contrary to predictions, costs impacted the intensity of punishments and rewards in similar ways: higher costs reduced the amount of response. From a subgame perfect equilibrium perspective, recipients should not reward or punish when doing so decreases their wealth. Although monetarily detrimental, over 80% in both the honest-relative and honest-absolute condition chose to reward. Only 35% of deceived recipients chose to punish. As a whole, the results suggest that recipients will be more likely to forgo economic benefits when reciprocating positively than

negatively, with the likelihood of rewards apparently impervious to increasing costs (at least in this case).

I also measured affective and cognitive responses to understand the psychological underpinnings of punishment and reward. The results support the literature on the norm of reciprocity (Gouldner, 1960; Gross & Latane, 1974; Kunz & Woolcott, 1976), the affective responses to negative behavior (Martorana, 2005; Pillutla & Murnighan, 1996; Sanfey et al., 2003), and the cognitive appraisals following positive behavior (King-Casas et al., 2005; Pillutla et al., 2003). In summary, (1) people felt more obligated to reward honesty than to punish deception, (2) trust mediated the relationship between honesty and reward, but not the relationship between deception and punishment, (3) negative affect mediated the relationship between deception and punishment, but not the relationship between honesty and reward, and (4) obligation mediated the relationship between trust and reward.

If reciprocal norms do indeed encourage recipients to reward honesty more than punish deception, then how do we reconcile the results with the overwhelming evidence supporting the negativity bias? Moreover, why is there such a strong assumption by researchers (c.f. Fehr & Gächter, 2000b) that people punish more than reward? Ironically, the rarity of punishments might make punishments more salient and more memorable – even for researchers – than rewards (Baumeister et al., 2001; Kellermann, 1984; Rozin & Royzman, 2001).

At first blush, stronger reactions to honesty than to deception might be a mere exception to the ‘bad is stronger than good’ literature. However, upon closer inspection, this finding may not necessarily conflict with but rather complement certain aspects of the bad-versus-good research. The bad is stronger than good literature suggests that individuals process negative

information more thoroughly than positive information (e.g., Anderson, 1965; Peeters & Czapinski, 1990; Skowronski & Carlston, 1989) because negative events occur less frequently than positive events. The greater frequency of positive events as compared to negative events may occur because many if not most of studies take place in prosperous societies (e.g., American and Western European societies) (Kellermann, 1984).

The rarity of punishments in comparison to rewards makes punishments more salient than rewards. In turn, individuals might assume that punishments occur more often than rewards because a negative action like punishing deception is more memorable than a positive act like rewarding honesty. Thus, we may not soon forget the downfall of Jeff Skilling and Bernie Ebbers, the CEOs of Enron and WorldCom, but will more quickly forget the names of Sherron Watkins and Cynthia Cooper, their respective whistle-blowers. All in all, deception might lead to stronger psychological reactions, but weaker behavioral responses – a worthwhile topic for future research.

Finally, Experiment 2 investigated whether affect and cognition influenced subsequent behavior. Results suggested that the same mediating variables of trust and negative affect impacted subsequent effects. The mediating factors impacted not only direct responses, but also potentially impacted subsequent behavior; trust reduced theft after honesty and trends suggested that negative affect may have increased theft. The practical implications of this finding will be discussed in greater detail in the general discussion.

CHAPTER 3

Cultural differences in rewarding honesty and punishing deception

Experiments 1 and 2 provided evidence that recipients rewarded honesty more frequently and intensely than they punished deception. However, the two experiments were run at a North American university where the majority of participants were of Western origin. Therefore, the experiments did not directly test whether responses generalized beyond Western culture. I was interested in the important moderating influence of culture (East Asian vs. Western) on reward and punishment behavior. Clearly understanding the psychological mechanisms motivating responses to deception and honesty within each culture may help managers develop techniques that encourage positive and cooperative behavior between members of different cultures.

Cultural differences in responses to deception and honesty might arise from fundamental differences in cognitive processing. Specifically, some have challenged the assumption of a universal pattern of thought and cognitive processing, questioning whether psychological processes tested within North America generalize to other cultures (A. P. Fiske, Kitayama, Markus, & Nisbett, 1998; Nisbett, 2003).

Differences between East Asian and Western dispositions are thought to arise from contrasting cultural historical traditions. East Asian cultures (e.g., Chinese, Japanese, and Korean cultures) are influenced by neo-Confucian and Buddhist teachings emphasizing social harmony while Western cultures (e.g., cultures of North European, Anglo-Saxon origin) are influenced by Judeo-Christianity emphasizing individual freedom. Based on these different value systems, I hypothesized that cultural differences in reciprocity norms and trust levels would drive these varying responses.

Felt obligation

The two historical cultural traditions have created two divergent views of morality (Chiu et al., 1997), each suggesting different levels of felt obligation as appropriate responses to deception and honesty. In the neo-Confucian tradition, each individual is seen as part of a larger microcosm that holds an implicit set of rules for moral and ethical conduct. A famous Japanese proverb, “The nail that sticks out is hammered down,” exemplifies this tradition. East Asian cultures are grounded in duty-based morality (Chiu & Hong, 1997; Hong, Ip, Chiu, Morris, & Menon, 2001), where meeting socially-established obligations is of utmost importance (Shweder & Miller, 1985). Duty-based moralists believe in the immutable nature of the world and tend to focus on whether others have carried out the duties prescribed by a rigid moral code. The end goal of duty-based moralists is to ensure social stability, with those deviating from the existing moral order punished for their wrongdoing. However, good behavior does not garner rewards because one is merely fulfilling a given duty (Chiu et al., 1997; Dworkin, 1978; Hamilton, Blumenfeld, Akoh, & Miura, 1990).

In contrast, the Judeo-Christian tradition views independence and individuality as most important (Hofstede, 1980; S. H. Schwarz, 1992). A quote by American author James Fenimore Cooper exemplifies this near-necessity for individuality, “All greatness of character is dependent on individuality. The man who has no other existence than that which he partakes in common with all around him, will never have any other than an existence of mediocrity.” Western cultures, influenced by Judeo-Christian values, are more grounded in rights-based morality (Chiu & Hong, 1997; Hong et al., 2001), where the focus is on protecting human rights. Rights-based moralists find sacrificing rights and individuality for social stability reprehensible. They are

more likely than duty-based moralists to view rewards as appropriate and acceptable and punishments as unnecessary and unwarranted incentives (Chiu et al., 1997).

Overall, Western culture suggests that the need for individual freedom overrides the East Asian belief in the need for social harmony. These different needs suggest that East Asians, as compared to Westerners, will feel more obligated to punish, and Westerners will be more inclined to reward.

Level of trust

Cultural differences in trust levels may also underlie contrasting reward behavior. Recall that trust was a key driver of reward behavior in Experiment 2; this implies that cultural differences in trust might significantly affect reward behavior. Recently, trust researchers have investigated trust levels in whether East Asians and Westerners (Buchan & Croson, 2004; Croson & Buchan, 1999; Fukuyama, 1995; Inglehart, Basanez, & Moreno, 1998; Yamagishi & Yamagishi, 1994; Yuki, Maddux, Brewer, & Takemura, 2005). Theorists have suggested that East Asians trust others less than Westerners do (Fukuyama, 1995; Inglehart et al., 1998; Yamagishi & Yamagishi, 1994). Moreover, research and theory suggest that trust levels between cultures differ depending on the closeness of the relationship between the trustee and trustor (Fukuyama, 1995; Inglehart et al., 1998; Yuki et al., 2005). As a whole, theorists suggest that Westerners exhibit more depersonalized trust – a type of trust in which personal relationships are not necessary or essential (Brewer, 1981) – than East Asians (Fukuyama, 1995; Inglehart et al., 1998; Yuki et al., 2005).

For example, Fukuyama (1995) has suggested that managers in the United States are more likely than Chinese managers to trust non-kin counterparts. He posits a boundary of trust:

those within the boundary are trusted and those outside the boundary are not. Westerners' boundaries are wider than East Asians, including people with whom they are more relationally distant (e.g., acquaintances). Evidence also suggests that just the potential for a future relationship will increase trust for East Asians more than for Westerners (Inglehart et al., 1998; Yuki et al., 2005).

Thus, this literature suggests that relationships affect the trust development of East Asians more than Westerners. Even when an actor performs a trustworthy action, East Asians may be more reticent than Westerners to trust that actor unless they expect future interactions. In other words, one act of kindness may influence the trust levels of East Asians far less than Westerners, especially with clear knowledge that there will be no future interaction between the two parties. As a result, one trustworthy action from a stranger may lead to greater rewards for strangers interacting with Westerners rather than East Asians.

Overview

Experiments 3 and 4 used the experimental designs in Experiments 1 and 2 to test for cultural differences in punishments for deception and rewards for honesty. These cultural differences may be attributed to divergent cultural philosophies that engender discrepant levels of felt obligation and trust. First, I predict that East Asians will feel more obligated to punish and less obligated to reward than Westerners. Second, I predict that East Asians will be less likely than Westerners to trust an honest actor, particularly when no future interaction is expected. The combination of these two predictions means that East Asians are likely to punish more and reward less than Westerners.

Experiment 3 compared the reactions to deception and honesty using the business scenario paradigm in Experiment 1. Study 4 reanalyzed data from Experiment 2, breaking the dataset into two cultural groups and comparing responses, felt obligation, and levels of trust of East Asians and Westerners. Both the Experiment 3 scenario and Study 4 instructions stressed the one-time nature of the interaction with the other party.

CHAPTER 4

Experiment 3: Cultural differences in scenario responses

In Experiment 3, Western participants responded to the business scenario described in Experiment 1. For the East Asian participants, the scenario was translated into traditional Chinese characters. Experiment 3 had a 2 (behavior of partner: honest-absolute, deception) x 2 (culture: East Asian, Western) x 2 (cost to respond: low, high) design.⁶

Method

Participants

Experiment 3 included a different set of participants from those in Experiment 1. Participants were 177 MBA students (106 men, 71 women) from Midwestern and Taiwan Universities. The Western subject pool included only participants of Anglo-Saxon descent born in the United States. The East Asian subject pool included participants of East Asian descent born in Taiwan. There were 76 East Asian participants and 101 Western participants. The questionnaires were given to participants as part of a class exercise.

Cross-Country Controls

To ensure that participants were relatively equivalent on an educational level, I compared MBA students from both Taiwan and the United States. This expanded the sample beyond the undergraduate students of Experiment 1. The instructions, scenario, and questions were given to Taiwanese students in both English and traditional Chinese characters to avoid any possible misinterpretation.

To ensure that the participants understood the monetary stake in terms of their own currency, the US dollar amounts were also presented in an equivalent amount of New Taiwan Dollars (NT), based on the exchange rate at the time the data was collected.

For example, in the dishonest condition, the scenario stated:

You and Pat recently completed a business deal and you have just found out that Pat was dishonest about some key information regarding the deal. As a result, you only received \$100 (approximately NT\$ 3300). You would have received 50% more if Pat had given you honest information.

The rest of the scenario remained the same as the original scenario. Please consult Appendix 5 for all of the translated questionnaires.

Dependent variables. As with Experiment 1 and 2, Experiment 3 measured the frequency and intensity, including and excluding those that had decided not to punish nor reward. I tested for gender effects in all of the analyses.

Results

Gender covariate

Similar to the previous experiments, I tested for gender differences in the Experiment 3 analyses. Gender was collapsed in all subsequent analyses because no significant interactions emerged, $F < 1$. I conducted two sets of analyses, one with and one without gender as a covariate, and the results did not differ significantly. Thus, the analyses reported did not include gender as a covariate.

Choice to respond (frequency)

I predicted that East Asians would punish more and reward less often than Westerners. The choice to respond was submitted to a 2 (behavior of partner: deception, honesty-absolute) x 2 (culture: East Asian, Western) x 2 (cost to respond: low, high) log-linear analysis. The behavior x culture x choice to respond interaction was not quite significant, $\chi^2(1) = 1.55, p = .21$. The frequencies suggested that East Asians punished more often (63%) than did Westerners (37%), and East Asians (74%) and Westerners (75%) did not differ in the frequency they rewarded honesty. The disproportionate number of participants per condition may have contributed to the non-significant effects (Figure 14 displays the frequencies and percentage of participants responding by condition).

Amount of response (including zero responses)

I predicted that East Asians would punish more and reward less than Westerners in terms of actual dollar amounts. Intensity was submitted to a 2 (behavior of business partner: deception, honesty-absolute) x 2 (culture: East Asian, Western) x 2 (cost to respond: low, high) between participants analysis of variance (ANOVA). There was a significant behavior x cost interaction, $F(1, 169) = 4.79, p < .05$. Most importantly, there was a significant behavior x culture interaction, $F(1, 169) = 5.92, p < .02$.

Planned contrasts revealed that East Asians ($M = 41.87, SD = 38.07$) punished deception significantly more than Westerners ($M = 20.29, SD = 30.73$), $F(1, 169) = 9.90, p < .01$. However, East Asians ($M = 35.71, SD = 28.99$) and Westerners ($M = 44.24, SD = 36.71$) did not statistically differ in the amounts they rewarded honesty, $F < 1$. In line with the past two experiments, Westerners rewarded honesty ($M = 44.24, SD = 36.71$) more than they punished deception ($M = 20.29, SD = 30.73$), $F(1, 169) = 6.32, p < .05$. However, East Asians punished

($M = 41.87$, $SD = 38.07$) more than they rewarded ($M = 35.71$, $SD = 28.99$), but not significantly more, $F(1, 171) = 1.01$, $p = .31$. The means by condition are shown in Figure 15.

Amount of response (excluding zero responses).

I excluded sixty-three participants who chose not to punish or reward. I analyzed the remaining data in a 2 (behavior of business partner: deception, honesty-absolute) x 2 (culture: East Asian, Western) x 2 (cost to respond: low, high) between participants ANOVA. The behavior x culture interaction was marginally significant, $F(1, 104) = 2.74$, $p = .10$. The planned contrasts revealed only one significant difference. East Asians punished deception ($M = 67.00$, $SD = 24.66$) more than they rewarded honesty ($M = 47.61$, $SD = 23.22$), $F(1, 104) = 9.94$, $p < .01$. Western responders rewarded ($M = 59.59$, $SD = 29.86$) and punished ($M = 54.62$, $SD = 25.37$) at similar levels, $F < 1$. Although the effects were not significant, East Asians ($M = 67.00$, $SD = 24.66$) punished more than Westerners ($M = 54.62$, $SD = 25.37$), $F(1, 104) = 2.59$, $p < .11$ (see Figure 16).

Discussion

Experiment 3 indicated that East Asians reported that they would punish more intensely than Westerners. Moreover, the data confirmed the reward-punishment asymmetries in Experiment 1 and 2, with Westerners rewarding more frequently and intensely than they punished. In contrast, East Asians did not differ in how they rewarded and punished when zero responses were included.

A limitation the scenarios was that the participants' business partner had an Americanized name, Pat. Given this information, East Asians may have punished more than Westerners because the American name may have suggested a cultural difference with the

business partner. As a result, East Asians may have viewed the business partner as an out-group member (Brewer & Brown, 1998; Brewer & Campbell, 1976) and punished based on out-group distinction rather than competing cultural philosophies. Experiment 4 addressed this limitation by making the identity of the actor anonymous and also explored cultural differences with actual rather than scenario-based responses.

CHAPTER 5

Experiment 2 reanalyzed by culture

The data from Experiment 2 was reanalyzed, testing for cultural differences in the responses of individuals of East Asian and Anglo-Saxon descent. The experiment had a 2 (behavior of partner: honest-absolute, deception) x 2 (culture: East Asian, Western) x 3 (cost to respond: no, low, high) design.

Method

Participants

The original dataset from Experiment 2 included 184 undergraduate students (83 males and 101 females) from a major Midwestern University. In the demographic questionnaire, participants reported their ethnicity by choosing one of six ethnic categories: Caucasian, African-American, Hispanic, East Asian, South Asian, or Other. For these analyses, only participants who checked the Caucasian or East Asian categories were included in the analyses. This left a total of 106 participants (54 males and 52 females; 75 Caucasians, 31 East Asians).

Dependent variables

The dependent variables were the same as those in Experiment 1-3: frequency and intensity (including and excluding zero responses), plus measures of trust, negative affect, and felt obligation. The scales ranged from 1-7, with higher numbers reflecting high levels of trust, negative affect, and felt obligation. Table 6 provides means and correlations for the entire sample on all of the variables reported in Experiment 2 reanalyzed.

Results

Gender covariate

Similar to the previous experiments, I tested for interactions with gender. Gender was collapsed in all subsequent analyses because no significant interactions emerged, $F < 1$. I conducted two sets of analyses, one with and one without gender as a covariate, and the results did not differ significantly. Thus, the analyses reported did not include gender as a covariate.

Choice to respond (frequency)

The choice to respond was analyzed in a logistic regression analysis. The explanatory terms included partner's behavior (categorical: deception, honesty-absolute), culture (categorical: Western, East Asian), and cost to respond (continuous), the interactions between the variables. The partner's behavior x culture interaction was the only significant interaction, $\chi^2(1, N = 106) = 4.56, p < .05$. The frequencies of responses are displayed in Figure 17. East Asians (69%) punished significantly more often than Westerners (26%), $\chi^2(1, N = 106) = 3.83, p < .05, \text{Exp}(B) = 11.17$. In contrast, East Asians (80%) did not statistically differ from Westerners (92%) in rewards, $\chi^2(1, N = 106) = 1.05, p = .30, \text{Exp}(B) = .32$.

I also compared punishment and reward frequencies within each culture. Westerners rewarded (92%) more than they punished (26%), $\chi^2(1, N = 106) = 23.03, p < .001, \text{Exp}(B) = 61.56$. However, East Asians punished (69%) and rewarded (80%) in approximately equal frequencies, $\chi^2(1, N = 106) < 1$.

Amount of response (including zero responses)

I analyzed intensity in a 2 (behavior of other player: deception, honesty-absolute) x 2 (culture: East Asian, Western) x 3 (cost to respond: no, low, high) between participants analysis of variance (ANOVA). The ANOVA yielded a significant interaction, $F(1, 94) = 13.98, p = .001$. Planned contrasts revealed that East Asians ($M = 1.78, SD = 1.67$) punished significantly

more than Westerners ($M = .68, SD = 1.27$), $F(1, 94) = 7.67, p < .01$. In contrast, East Asians rewarded significantly less ($M = 1.67, SD = 1.32$) than Westerners ($M = 2.65, SD = 1.38$), $F(1, 94) = 4.82, p < .05$. In line with the results from Experiments 1-3, Westerners rewarded honesty more than they punished deception, $F(1, 94) = 38.60, p < .001$, whereas East Asians rewarded and punished at approximately similar amounts, $F < 1$ (see Figure 18 for the means and frequencies by condition).

Amount of response (excluding zero responses)

Thirty-eight participants did not respond, leaving 68 participants who did. I analyzed intensity in a 2 (behavior of other player: deception, honesty-absolute) x 2 (culture: East Asian, Western) x 3 (cost to respond: no, low, high) between participants ANOVA. The behavior of partner x culture interaction was suggestive, but not significant, $F(1, 57) = 2.36, p = .13$. The results suggested that Westerners ($M = 2.87, SD = 1.19$) rewarded more than East Asians ($M = 2.08, SD = 1.12$), whereas East Asians ($M = 2.59, SD = 1.38$) and Westerners ($M = 2.50, SD = 1.15$) punished at approximately similar amounts (Figure 19).

Level of Trust

Trust levels were analyzed in a 2 (behavior of partner: deception, honesty) x 2 (culture: East Asian, Western) x 3 (cost to respond: no, low, high) between participants ANOVA (see Figure 19). The only significant interaction was behavior x culture, $F(1, 94) = 4.94, p < .05$. As expected, a significant main effect indicated that recipients trusted honest actors ($M = 5.33, SD = .82$) more than dishonest actors ($M = 3.33, SD = 1.00$), $F(1, 94) = 9.21, p < .01$. Planned contrasts revealed that after honesty, East Asians ($M = 4.95, SD = .71$) trusted less than Westerners did ($M = 5.50, SD = .82$), $F(1, 94) = 4.19, p = .05$. After deception, East Asians (M

= 3.55, $SD = .92$) and Westerners ($M = 3.23$, $SD = 1.02$) trusted the actors at similar levels, $F(1, 94) = 1.18$, $p = .28$ (see Figure 19). Westerners trusted honest ($M = 5.50$, $SD = .82$) more than dishonest actors, ($M = 3.23$, $SD = 1.02$), $F(1, 94) = 109.31$, $p < .001$. Likewise, East Asians trusted honest ($M = 4.95$, $SD = .71$) more than dishonest actors, ($M = 3.55$, $SD = .92$), $F(1, 94) = 17.53$, $p < .001$ (see Figure 20 for the means and frequencies by condition).

Negative affect

Negative affect was analyzed in a 2 (behavior of partner: deception, honesty) x 2 (culture: East Asian, Western) x 3 (cost to respond: no, low, high) between participants ANOVA (see Figure 19). The only significant interaction was behavior x culture, $F(1, 94) = 4.94$, $p < .05$ (see Figure 21). As expected, a significant main effect indicated that recipients were unhappier with dishonest actors ($M = 3.96$, $SD = .87$) than with honest actors ($M = 2.97$, $SD = 1.00$), $F(1, 94) = 16.14$, $p < .001$. Planned contrasts revealed that after honesty, East Asians ($M = 3.93$, $SD = .51$) unhappier than Westerners ($M = 3.13$, $SD = 1.39$), $F(1, 94) = 4.21$, $p < .05$. After deception, East Asians ($M = 3.80$, $SD = .55$) and Westerners ($M = 4.02$, $SD = .92$) exhibited similar levels of unhappiness, $F < 1$. The results also suggested that Westerners were unhappier after dishonesty than honesty, $F(1, 94) = 30.55$, $p < .001$, however, East Asians exhibited the same level of affect after honesty and dishonesty, $F(1, 94) = 1.40$, $p = .24$.

Felt Obligation

Felt obligation was analyzed in a 2 (behavior of partner: deception, honesty) x 2 (culture: East Asian, Western) x 3 (cost: no, low, high) between participants ANOVA. Only the 3-way interaction was significant, $F(1, 94) = 2.21$, $p = .05$. The results suggested that particularly at high costs, East Asians felt more obligated to punish deception ($M = 3.50$; $SD = 2.35$) than did

Westerners ($M = 1.93$; $SD = 1.16$), $F(1, 94) = 3.79$, $p = .05$. Moreover, the data suggested that Westerners felt more obligated to reward ($M = 3.89$; $SD = 1.67$) than to punish ($M = 2.78$; $SD = 1.83$), $F(1, 94) = 6.67$, $p = .01$, and that East Asians did not differ in felt obligation to reward ($M = 3.40$; $SD = 1.35$) and punish ($M = 3.44$; $SD = 1.89$), $F(1, 94) = 1.10$, $p = .30$ (see Figure 22).

Discussion

The results from Experiment 4 suggested that East Asians punished more frequently and intensely than did Westerners. East Asians also rewarded about as frequently as but less intensely than Westerners. No cultural differences in punishments and rewards emerged when zero-response participants were omitted.

When examining within-culture responses, an interesting pattern emerged: Westerners consistently rewarded more than they punished, whereas East Asians did not differ in punishment and reward frequencies and intensities. Implications for these findings will be discussed in greater detail in the next section.

Experiment 4 also shed some light on the psychological mechanisms underlying responses. The data – not statistically significant – suggested that East Asians felt more obligated to punish and less obligated to reward than Westerners. Also, as predicted, East Asians trusted honest actors less than did Westerners. East Asians and Westerners did not exhibit different levels of trust for dishonest actors. These findings support the theoretical argument that cultural differences in felt obligation and trust can engender divergent punishment and reward behavior by East Asians and Westerners.

Interestingly, over 40% of the East Asian participants in Experiment 4 were born and raised in the United States. This leaves open the question of whether differences might be even greater for native East Asians.

CHAPTER 6

GENERAL DISCUSSION

My dissertation compared how recipients rewarded honesty and punished deception in these experiments. Participants in Experiments 1 and 3 read scenarios about a deceptive or honest business partner and allocated hypothetical punishments and rewards. Participants in Experiment 2 responded to deception and honesty within a laboratory setting with actual financial incentives and consequences. In Experiments 1 and 2, contrary to the previous findings, recipients rewarded honesty more frequently and intensely than they punished deception (Abbink et al., 2000; Brandts & Charness, 2003; Engelmann & Ortmann, 2001; Offerman, 2002; Pereira et al., 2006). Experiments 3 and Experiment 2 re-analyzed by culture indicated that responses to deception and honesty depend, at least in part, on cultural forces: Whereas Westerners rewarded more frequently and intensely than they punished, East Asians punished and rewarded at approximately equivalent frequencies and intensities. Cross-cultural comparisons from both experiments also revealed that East Asians punished more frequently and intensely, while rewarding less frequently than Westerners. Data from Experiment 2 re-analyzed also suggested that East Asians rewarded less intensely than Westerners.

All three experiments explored how costs moderated the frequencies and amounts of punishment and reward. Experiments 1 and 3 compared low and high costs to respond; Experiment 2 compared no, low, and high costs to respond. The scenario data indicated that costs influenced the frequency and size of punishments and rewards. However, data from Experiment 2 suggested that as costs increased, the likelihood to punish decreased, whereas the likelihood to reward remained constant, with a large majority of participants (>75%) choosing to

reward even when it was costly to do so. This finding suggests that more people maybe utility-maximizing when punishing (e.g., did not punish when it was costly) than when rewarding.

Psychological Underpinnings

Post hoc analyses indicated that, in the context of one-shot interactions, people, specifically Westerners, rewarded more than they punished because they felt more obligated to reciprocate kind behavior than hurtful behavior (Cialdini, 1984; Eisenberger et al., 1986; Pillutla et al., 2003; Tesser et al., 1968). Specifically, the data from Experiment 2 suggested that Westerners felt more obligated to reward honesty than to punish deception; East Asians, in contrast, felt equally obligated to punish and reward. These results support the idea that divergent responses may originate from different moral traditions (Chiu et al., 1997), i.e. recipients from duty-based East Asian societies punished more and rewarded less than those from individually oriented rights-based Western societies.

Moreover, Experiment 2 indicated that Western responses were driven by two distinct psychological systems, with affective responses (negative affect) to deception influencing punishment intensity and cognitive appraisals (perceptions of trustworthiness) following honesty influencing felt obligation to reward and reward intensity. Experiment 2 re-analyzed suggested that cultural differences in trust perceptions directly influenced reward amounts, with East Asians trusting and rewarding other players less than did Westerners. Like Nisbett (2003), these experiments suggest that Westerners' cognitive and behavioral responses do not necessarily generalize to all cultures.

These results run counter to the predictions highlighting the strong impact of negatives over positives (Baumeister et al., 2001; Kube, Marechel, & Puppe, 2007; Rozin & Royzman,

2001). Ironically, negativity itself may contribute to the prediction that individuals punish more than reward. One theoretical rationale is that negative stimuli are more salient and weighted more heavily than positive stimuli because they depart more from what is normative (Kellermann, 1984). Thus, in Western cultures, punishments may be considerably more salient and memorable than rewards because punishments are rarer than rewards.

In addition, the majority of evidence supporting the negativity bias focuses on impression formation and person perception rather than behavioral responses (Baumeister et al., 2001; Kellermann, 1984). For example, a deceptive individual is disliked more than an honest individual is liked (Falk & Fischbacher, 2006) and negative behavior is more often attributed to one's disposition than is positive behavior (Skowronski & Carlston, 1989). The current experiments suggest that these perceptual tendencies may not necessarily trigger greater punishments than rewards. Rather, when it comes to behavior, the results suggest that people reciprocate positively more than often negatively. This would suggest a model that differentially weights the impact of negatives on cognition (big) and their impact on behavior (small).

The fact that perception often precedes behavior (Bargh, 1997; Bargh, Chen, & Burrows, 1996; Bargh & Ferguson, 2000; Dijksterhuis & Bargh, 2001; Dijksterhuis & van Knippenberg, 1998) means that something must intervene to limit the strong impact of negatives on perceptions and judgment in their translation, or lack of translation, to behavior. One possible explanation is that the lack of punishment may arise from recipients avoiding rather than directly approaching and punishing a deceptive target (Elliot, 1999; W. James, 1890). Since the presence of social contact is fundamental to well-being (Baumeister & Leary, 1995), avoidant behavior and social exclusion may also be used as form of indirect punishment in Western cultures. A

lack of social connectedness produces deleterious repercussions (Baumeister & Leary, 1995), including high levels of stress hormones (Uchino, Cacioppo, & Kiecolt-Glazer, 1996) and impaired cognitive functioning (Williams, 2001).

In the United States, social exclusion has been used as a form of official punishment. At an individual level, prison inmates who have committed particularly dangerous crimes have been sentenced to solitary confinement, where they are denied contact with other human beings. Increased social isolation during solitary confinement has been associated with severe psychiatric symptoms (Grassian, 1983). At a national level, economic sanctions have become an increasingly popular tool in U.S. foreign policy. Experiments have shown that sanctions provide a useful tool in encouraging cooperation (Falk & Fehr, 2005). Future research, then, might compare when recipients use socially exclusion versus direct retaliation to punish.

The finding that East Asians punished more often than did Westerners suggests that they may be less likely to use avoidance as a form of punishment: a recent discussion by Kitayama, Markus, & Kurokawa (2000) is consistent with this argument. The authors suggest that an independent view of the self (Westerners) increases the motivation to maintain positive internal attributions by focusing and elaborating on positive feelings while decreasing and avoiding negative feelings. Those with an interdependent perspective (East Asians) consider internal attributions to be less important than the social aspects of the self (Cousins, 1989; A. P. Fiske et al., 1998), placing less weight on internal states and more on social engagement.

The viewpoint that East Asians will engage in behaviors that involve interaction, regardless of valence, is also consistent with the current findings that East Asians may punish and reward at similar frequencies. This raises a question about the general applicability of the

negativity bias (Baumeister et al., 2001; Kellermann, 1984); as a whole, the bad-good axis may be considerably less relevant for an East Asian population than for a Western population. Future research could pursue this further.

Practical Contributions

Organizational Implications

Not only does this dissertation encourage a synthesis of the economic and social psychological literatures and provide theoretical insight to the comparative responses to deception versus honesty, it also provides prescriptive implications for understanding interpersonal conflicts and disputes by providing potential insight into the failures to punish corporate wrongdoing. Whereas many studies on dispute resolution focus on how to reduce conflict between individuals, conflict in certain situations may benefit organizational functioning and interpersonal relationships. Punishing deception, rather than not responding at all, might be extremely important in deterring continued deception. If left unchecked, seemingly inconsequential white lies might eventually spiral into greater levels of deception within an organization, resulting in hazardous situations such as those seen at Enron and Tyco. “Star performers that violate the company's procedures are too often given a second chance,” said Sherron Watkins, the Enron whistle-blower, perfectly summarizing the pitfall of not adequately punishing deception.

The current results suggest that employees might be reticent to blow the whistle when the costs to respond are high. Recent legislation has attempted to counteract some of these deterrents. In 1986, Senator Charles Grassley and Representative Howard Berman passed a revised False Claims Act that allowed whistleblowers who launched litigation to receive up to 25

percent of any money recovered by their lawsuit, thereby allaying some of the costs to punish deception.

However, encouraging punishment, especially the negative emotions associated with punishment, should be taken with a grain of salt. The current evidence found that, in line with past research (Mullen & Nadler, 2007), people who were angrier over deception were more likely to take a borrowed pen from the experimenter. The irony, therefore, is that the negative emotions that are associated with the constructive punishment of deception may also promote harmful behaviors such as theft.

In contrast, the trust that arises from honesty engenders both direct *and* indirect positive behaviors. Greater trust in an honest partner not only increased rewards, but also decreased the likelihood of stealing. Not only incidental emotions (Forgas, 1995; Loewenstein, 1996; N. Schwarz, 1990), but cognitions from one situation influenced normatively unrelated decisions. Thus, a vital undertaking for managers is to understand how to encourage the punishment of deception without the associated negative emotions, while encouraging behaviors that transmit trust within the organization.

Cross-cultural Implications

Considering how recipients reward honesty and punish deception across different cultures may also be vital for organizations that employ increasingly international workforces. Specifically, managers may be able to improve group and organizational performance if they clearly understood how incentives affected coworkers from different cultural backgrounds. If East Asians punish more than Westerners in various contexts (Friedman et al., 2007; Kelley & Tseng, 1992), East Asians, compared to Westerners, may be more receptive to punishments as a

form of deterrent. East Asians may be more accepting of punishments because happiness is derived from social harmony rather than from personal happiness (Christopher, 1999; Kitayama et al., 2000).

Clashes of culture and expectation might also be predictable. One of the most prominent examples of cultures colliding over the acceptability of punishment is the 1994 caning of Michael Fay, a United States teenager who was charged with vandalizing public property in Singapore, a Southeast Asian city-state with a Chinese ethnic majority. The ruling sparked outrage from the public and the United States government, culminating in a statement from U.S. President Bill Clinton who called the punishment excessive and requested clemency. Many Singaporeans were disgruntled with the United States' attempt to intervene, noting that Singapore, as a sovereign state, could use its own discretion in extending punishments. Moreover, the harsh penal system was viewed as a community-oriented system that provided an extremely safe environment for Singaporean citizens (Chew, 1994). This example clearly exemplifies divergent moral perspectives of the two different cultures: the United States viewed the severity of punishment as a moral outrage, overstepping individual rights, whereas Singapore viewed the punishment as necessary to maintain social solidarity.

Conclusion

As a whole, the data in this dissertation provided strong support for a model that describes likely responses to deception and honesty (see Figure 1). Together, the experiments gave insight into key mediating and moderating variables driving both direct and indirect responses to deception and honesty, and the results challenged the empirical findings that people punish deception more than they reward honesty (Abbink et al., 2000; Brandts & Charness,

2003; Engelmann & Ortmann, 2001; Offerman, 2002; Pereira et al., 2006). Several questions remained open about these findings because some of the experiments suffered from asymmetric study designs (e.g., higher costs to punish than reward). Thus, the current experiments attempted a more thorough, systematic approach comparing the frequencies and intensities of punishment and reward while holding levels of deception and honesty constant and equivalent. Moreover, two different contexts and operationalizations increased the scope and the potential validity of the results.

In conclusion, the dynamics of punishment and reward may not be mirror images of one another. Rather, they may be activated by different cognitive and/or affective mechanisms, which ultimately lead to different behavioral choices. This line of research also provides a distinct perspective on the ‘bad is greater than good’ argument, taking it in a new direction that warrants greater theoretical reconceptualization. Continuing to chart these connections provides for an intriguing line of research and study.

ENDNOTES

¹ The initial data collection of Experiments 1 and 2 only included the deception and honesty-absolute conditions. The honesty-relative condition was collected at a later date to assuage reviewers' concerns and to make the results more comprehensive. A thorough analysis ensured that no demographic differences (e.g., gender, age, and ethnicity) existed between the first and second data collections in both experiments.

² Intensity and frequency of response did not differ by whether or not participants believed the other player. Therefore, this factor was collapsed in all subsequent analyses in Experiment 2.

³ An anger measure was also included in the post-questionnaire (e.g., 1. After you found out the information about the other player's choice in Stage 1, to what degree did you feel the following emotion? ANGRY). While anger is cited as the precursor to retaliation behavior (Martorana, 2005; Pillutla & Murnighan, 1996), the anger measure was not associated with intensity of response or pen theft. Because anger may have dissipated by the post-questionnaire, future research should directly manipulate emotions such as anger and unhappiness to clearly understand the specific negative emotions associated with punishment behavior.

⁴ Planned chi-squared tests revealed that in the honesty-absolute condition, the no cost (93%), low cost (77%) and high cost (90%) conditions did not differ, $\chi^2(1) < 2.5$. Similarly, in the honesty-relative condition, the frequency of reward did not statistically differ between the no cost (100%), low cost (93%), and high cost (83%) conditions. Except for the comparison between the no cost and high cost condition, $\chi^2(1) = 1.83, p = .18$, the two other comparisons

yielded F 's < 1 . These results suggest that recipients of honesty were unaffected by cost; rewards were similar regardless of costs.

⁵ In their classic paper on mediation, Baron and Kenny (1986) required the total effect be significant prior to any tests of mediation. However, more recently researchers have suggested that this step be relaxed. In particular, Shrout and Bolger (2002) advocate that the direct relationship between the independent variable and dependent variable need not be significant prior to testing indirect effects because they might be suppressed by a competing process (see Shrout & Bolger, 2002, for an excellent discussion on suppression).

⁶ I am currently collecting Taiwanese data for the honest-relative condition in Experiment 3. I excluded the honesty-absolute condition in Experiment 2 reanalyzed because it only contained a small subset of East Asians; data is being collected to alleviate this limitation.

REFERENCES

- Abbink, K., Irlenbusch, B., & Renner, E. (2000). The moonlighting game: An experimental study on reciprocity and retribution. *Journal of Economic Behavior and Organization*, *42*, 265-277.
- Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. 1965. *Journal of Experimental Psychology*, *70*(4), 394-400.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life: Advances in social cognition, Vol. 10. Advances in social cognition, Vol. 10* (pp. 1-61). Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230-244.
- Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*(6), 925-945.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323-370.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497-529.

- Bazerman, M. H., Magliozzi, T., & Neale, M. A. (1985). Integrative bargaining in a competitive market. *Organizational Behavior & Human Decision Processes*, 35(3), 294-313.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social-History. *Games and Economic Behavior*, 10(1), 122-142.
- Blau, P. M. (1964). *Exchange and power in social life*. New York: John Wiley.
- Boles, T. L., Croson, R. T. A., & Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior & Human Decision Processes*, 83(2), 235-259.
- Bosman, R., Sutter, M., & van Winden, F. (2005). The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology*, 26(3), 407-429.
- Brandts, J., & Charness, G. (2003). Truth or consequences: An experiment. *Management Science*, 49(1), 116-130.
- Brewer, M. B. (1981). Ethnocentrism and its role in interpersonal trust. In M. B. Brewer & B. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 345-360). San Francisco, CA: Jossey-Bass.
- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, 55(3), 429-444.
- Brewer, M. B., & Brown, R. J. (1998). Intergroup relations. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Series Ed.) & *The handbook of social psychology* (4 ed., pp. 554-594). Boston: McGraw-Hill.
- Brewer, M. B., & Campbell, D. T. (1976). *Ethnocentrism and intergroup attitudes: East African evidence*. Beverly Hills, CA: Sage.

- Brosig, J., Weimann, J., & Yang, C.-L. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics*, 6, 75-90.
- Buchan, N., & Croson, R. (2004). The boundaries of trust: own and others' actions in the US and China. *Journal of Economic Behavior & Organization*, 55(4), 485-504.
- Cacioppo, J. T., & Gardner, W. L. (1999). Emotions. *Annual Review of Psychology*, 50, 191-214.
- Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, 76(5), 839-855.
- Camerer, C. (2003). *Behavioral Game Theory*. Princeton: Princeton University Press.
- Camerer, C., & Thaler, R. H. (1995). Ultimatums, Dictators and Manners. *Journal of Economic Perspectives*, 9(2), 209-219.
- Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry*, 37(1), 47-59.
- Cannon, L. (1999). *Official Negligence: How Rodney King and the Riots Changed Los Angeles and the LAPD*. Boulder, CO: Westview.
- Cesario, J. E., Plaks, J. E., & Higgins, E. T. (in press). Automatic social behavior as motivated preparation to interact. *Journal of Personality and Social Psychology*.
- Chaudhuri, A., & Gangadharan, L. (2007). Gender Differences in Trust and Reciprocity. *Unpublished Manuscript*.
- Chew, M. (1994). Human-Rights in Singapore - Perceptions and Problems. *Asian Survey*, 34(11), 933-948.

- Chiu, C.-Y., Dweck, C. S., Tong, J. Y.-Y., & Fu, J. H.-Y. (1997). Implicit theories and conceptions of morality. *Journal of Personality and Social Psychology, 73*(5), 923-940.
- Chiu, C.-Y., & Hong, Y. (1997). Justice from a Chinese perspective. In H. S. R. Kao & D. Sinha (Eds.), *Asian perspectives on psychology* (p. 164–184). Thousand Oaks, CA: Sage.
- Christopher, J. C. (1999). Situating psychological well-being: Exploring the cultural roots of its theory and research. *Journal of Counseling & Development, 77*(2), 141-152.
- Cialdini, R. B. (1984). The Triggers of Influence. *Psychology Today, 18*(2), 40-45.
- Cialdini, R. B., Green, B. L., & Rusch, A. J. (1992). When tactical pronouncements of change become real change: The case of reciprocal persuasion. *Journal of Personality and Social Psychology, 63*(1), 30-40.
- Cousins, S. (1989). Culture and selfhood in Japan and the U.S. *Journal of Personality and Social Psychology, 56*, 124-131.
- Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An interdisciplinary review. *Journal of Management, 31*(6), 874-900.
- Croson, R., & Buchan, N. (1999). Gender and culture: International experimental evidence from trust games. *American Economic Review, 89*(2), 386-391.
- Dijksterhuis, A., Aarts, H., Bargh, J. A., & van Knippenberg, A. (2000). On the relation between associative strength and automatic behavior. *Journal of Experimental Social Psychology, 36*(5), 531-544.
- Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol. 33* (pp. 1-40). San Diego, CA, US: Academic Press, Inc.

- Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology*, 74(4), 865-877.
- Duffy, M. K., Ganster, D. C., & Pagon, M. (2002). Social undermining in the workplace. *Academy of Management Journal*, 45(2), 331-351.
- Dufwenberg, M., & Kirchsteiger, G. (2000). Reciprocity and wage undercutting. *European Economic Review*, 44(4-6), 1069-1078.
- Dworkin, R. (1978). *Taking rights seriously*. Cambridge, MA: Harvard University Press.
- Eisenberger, R., Armeli, S., Rexwinkel, B., Lynch, P. D., & Rhoades, L. (2001). Reciprocation of perceived organizational support. *Journal of Applied Psychology*, 86(1), 42-51.
- Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology*, 71(3), 500-507.
- Eisenberger, R., Lynch, P., Aselage, J., & Rohdieck, S. (2004). Who Takes the Most Revenge? Individual Differences in Negative Reciprocity Norm Endorsement. 2004. *Personality and Social Psychology Bulletin*, 30(6), 789-799.
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169-189.
- Engelmann, D., & Ortmann, A. (2001). The Robustness of Laboratory Gift Exchange: A Reconsideration. *Unpublished Manuscript*.
- Exline, J. J., & Baumeister, R. F. (2007). Forgiveness in prisoner's dilemma games. *Unpublished findings, Case Western Reserve University*.

- Falk, A., & Fehr, E. F. U. (2005). Driving Forces Behind Informal Sanctions. *Econometrica*, 73(6), 2017-2030.
- Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games & Economic Behavior*, 54(2), 293-315.
- Fehr, E., & Gächter, S. (2000a). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980-994.
- Fehr, E., & Gächter, S. (2000b). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3), 159-181.
- Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does Fairness Prevent Market Clearing - an Experimental Investigation. *Quarterly Journal of Economics*, 108(2), 437-459.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3), 52.
- Fehr, E., Tougareva, E., & Fischbacher, U. (2002). Do high stakes and competition undermine fairness? *Unpublished Manuscript*.
- Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 915-981). New York: McGraw-Hill.
- Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Experimental Research in Personality*, 22, 889-906.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York, NY, England: McGraw-Hill Book Company.

- Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, 117(1), 39-66.
- Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6(3), 347-369.
- Friedman, R., Liu, W., Chen, C., & Chi, S.-C. (2007). Chinese and American Arbitrators: Examining the Effects of Attributions and Culture on Award Decisions. *Unpublished Manuscript*.
- Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. New York, NY: The Free Press.
- Gneezy, U. (2005). Deception: The Role of Consequences. *American Economic Review*, 95(1), 384-394.
- Gouldner, A. W. (1960). The Norm of Reciprocity - a Preliminary Statement. *American Sociological Review*, 25(2), 161-178.
- Grassian, S. (1983). Psychopathological Effects of Solitary Confinement. *American Journal of Psychiatry*, 140(11), 1450-1454.
- Greenberg, M. S. (1980). A theory of indebtedness. In K. Gergen, M. S. Greenberg & R. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 77-94). New York: Plenum.
- Gross, A. E., & Latane, J. G. (1974). Receiving help, reciprocation, and interpersonal attraction. *Journal of Applied Social Psychology*, 4(3), 210-223.
- Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behaviour and Organization*, 3, 367-388.

- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834.
- Hamilton, V. L., Blumenfeld, P. C., Akoh, H., & Miura, K. (1990). Credit and blame among American and Japanese children: Normative, cultural, and individual differences. *Journal of Personality and Social Psychology*, *59*(3), 442-451.
- Hauser, M. D. (2006). *Moral minds: how nature designed our universal sense of right and wrong*. New York: Ecco.
- Helm, B., Bonoma, T. V., & Tedeschi, J. T. (1972). Reciprocity for harm done. *Journal of Social Psychology*, *87*(1), 89-98.
- Hofstede, G. (1980). *Culture's consequences*. Beverly Hills, CA: Sage.
- Hokanson, J. E. (1974). An escape-avoidance view of catharsis. *Criminal Justice and Behavior*, *1*, 195-223.
- Holmbeck, G. N. (2002). Post-hoc probing of significant moderational and mediational effects in studies of pediatric populations. *Journal of Pediatric Psychology*, *27*(1), 87-96.
- Hong, Y.-y., Ip, G., Chiu, C.-y., Morris, M. W., & Menon, T. (2001). Cultural identity and dynamic construction of the self: collective duties and individual rights in Chinese and American cultures. *Social Cognition*, *19*(3), 251-268.
- Inglehart, R., Basanez, M., & Moreno, A. (1998). *Human cultures and beliefs: A cross-cultural sourcebook: Political, religious, sexual, and economic norms in 43 societies: Findings from the 1990-1993 world values survey*. Ann Arbor: University of Michigan Press.
- James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, *69*(2), 307-321.

- James, W. (1890). *The principles of psychology* (Vol. 2). New York: Holt.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the Assumptions of Economics. *Journal of Business*, 59(4), S285-S300.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory - Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291.
- Kellermann, K. (1984). The Negativity Effect and Its Implications for Initial Interaction. *Communication Monographs*, 51(1), 37-55.
- Kelley, M. L., & Tseng, H. M. (1992). Cultural-Differences in Child-Rearing - a Comparison of Immigrant Chinese and Caucasian American Mothers. *Journal of Cross-Cultural Psychology*, 23(4), 444-455.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, 308(5718), 78-83.
- Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, Emotion, and Well-being: Good Feelings in Japan and the United States. *Cognition and Emotion*, 14(1), 93-124.
- Kube, S., Marechel, M. A., & Puppe, C. (2007). Putting Reciprocity to Work - Positive versus Negative Responses in the Field. *Unpublished Manuscript*.
- Kunz, P. R., & Woolcott, M. (1976). Season's greetings: From my status to yours. *Social Science Research*, 5(3), 269-278.
- Leventhal, G. S., & Anderson, D. (1970). Self-interest and the maintenance of equity. *Journal of Personality and Social Psychology*, 15(1), 57-62.

- Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65(3), 272-292.
- Martorana, P. (2005). From System Justification to System Condemnation: Antecedents of Attempts to Change Status Hierarchies. *Unpublished Dissertation*.
- McAllister, D. J. (1995). Affect-Based and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal*, 38(1), 24-59.
- McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1996). Game theory and reciprocity in some extensive form experimental games. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23), 13421-13428.
- McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1998). Reciprocity, trust, and payoff privacy in extensive form bargaining. *Games and Economic Behavior*, 24(1-2), 10-24.
- 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.
- Moore, G. E. (1903). *Principia Ethica* (C. U. Press, Ed.). Cambridge.
- Mullen, E., & Nadler, J. (2007). Moral Spillovers: The Effect of Moral Mandate Violations on Deviant Behavior. *Unpublished Manuscript*.
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, 89(6), 852-863.
- Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. New York: Free Press.

- Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, 46(8), 1423-1437.
- Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, Games, and Common-Pool Resources*. Ann Arbor: University of Michigan Press.
- Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 1, pp. 33-60). New York: Wiley.
- Pereira, P. T., Silva, N., & Silva, J. (2006). Positive and Negative Reciprocity in the Labour Market. *Journal of Economic Behavior and Organization*, 59, 406-422.
- Pillutla, M. M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, 39(5), 448-455.
- Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior & Human Decision Processes*, 68(3), 17.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83(5), 22.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296-320.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626), 1755-1758.

- Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 527-561). New York: Guilford Press.
- Schwarz, S. H. (1992). Universals in the structure and content of values. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1-65). Orlando, FL: Academic Press.
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55(1), 5-14.
- Selten, R. (1965). Spieltheoretische behandlung eines oligopol modells mit nachfragefragheit. *Zeitschrift fur Gesamte Staatswissenschaft*, 121, 301–324.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and non-experimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422-445.
- Shweder, R. A., & Miller, J. G. (1985). The social construction of the person: How is it possible? In K. J. Kenneth & K. E. Davis (Eds.) (Ed.), *The social construction of the person* (p. 41–69). New York: Springer-Verlag.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131-142.
- Snijders, C., & Keren, G. (2001). Do you trust? Whom do you trust? When do you trust? In S. R. Thye, E. J. Lawler, M. W. Macy & H. A. Walker (Eds.), *Advances in Group Processes* (Vol. 18). Amsterdam: JAI, Elsevier Science.

- Staub, E. (1972). The use of role playing and induction in children's learning of helping and shaping behavior. *Child Development, 42*, 805-816.
- Tepper, B. J. (2000). Consequences of abusive supervision. *Academy of Management Journal, 43*(2), 178-190.
- Tesser, A., Gatewood, R., & Driver, M. (1968). Some Determinants of Gratitude. *Journal of Personality and Social Psychology, 9*(3), 233-236.
- Thaler, R. (1980). Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior & Organization, 1*(1), 39-60.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.
- Uchino, B. N., Cacioppo, J. T., & Kiecolt-Glazer, J. K. (1996). The relationship between social support and physiological processes: A review with emphasis on underlying mechanisms and implications for health. *Psychological Bulletin, 119*, 488-531.
- Uhl-Bien, M., & Maslyn, J. M. (2003). Reciprocity in manager-subordinate relationships: Components, configurations, and outcomes. *Journal of Management, 29*(4), 511-532.
- Wang, C. S., & Thompson, L. L. (2006). The negative and positive psychology of leadership and group research. In S. R. Thye & E. J. Lawler (Eds.), *Advances in Group Processes: Social psychology of the workplace* (Vol. 23, pp. 31-61). Amsterdam: JAI Press.
- Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*(6), 797-826.

- Williams, K. D. (2001). *Ostracism: The power of silence*. New York: Guilford.
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*, 129-166.
- Youngs, G. A. (1986). Patterns of threat and punishment reciprocity in a conflict setting. *Journal of Personality and Social Psychology, 51*(3), 541-546.
- Yuki, M., Maddux, W. W., Brewer, M. B., & Takemura, K. (2005). Cross-Cultural Differences in Relationship- and Group-Based Trust. *Personality and Social Psychology Bulletin, 31*(1), 48-62.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*(2), 151-175.
- Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist, 39*(2), 117-123.
- Zola-Morgan, S., Squire, L. R., Alvarez-Royo, P., & Clower, R. P. (1991). Independence of memory functions and emotional behavior: separate contributions of the hippocampal formation and the amygdala. *Hippocampus, 2*(1), 207-220.

TABLES

Table 1. Gender frequencies by experimental condition, Experiment 1

		Recipient's Action					
		Punish		Reward-Absolute		Reward-Relative	
		Low	High	Low	High	Low	High
Gender	Male	11	3	5	9	6	10
	Female	8	11	13	9	13	6

Table 2. Gender frequencies by experimental condition, Experiment 2

		Recipient's Action								
		Punish			Honesty-Absolute			Honesty-Relative		
Gender	Cost	No	Low	High	No	Low	High	No	Low	High
		Male	13	6	13	15	9	11	2	7
	Female	7	14	16	13	13	9	8	8	9

Table 3. Means, standard deviations, and correlations for independent and dependent variables, Experiment 2 (N = 176)

	Mean	<i>SD</i>	1	2	3	4
1. Partner's behavior (0 = D, 1 = H)	.61	.49	1			
2. Cost	1.02	.82	-.11	1		
3. Amount of response (intensity)	1.72	1.54	.42***	-.30***	1	
4. Choice to respond (frequency)	.68	.47	.56***	-.24**	.78***	1
5. Pen theft	.23	.42	.10	-.16*	.11	.14*
6. Felt obligation	3.34	1.76	.22**	-.10	.46***	.44***
7. Negative affect	4.65	1.06	-.51***	-.26**	-.32***	-.32***
8. Trust	4.54	1.32	.71***	-.12	.35***	.46***
9. Gender (0 = M, 1 = F)	.55	.50	.02	.06	.14 ⁺	.13 ⁺

⁺ p < .10; * p < .05; ** p < .01; *** p < .001

Table 3 continued. Means, standard deviations, and correlations for independent and dependent variables, Experiment 2 (N = 176)

	5	6	7	8	9
1. Partner's behavior (0 = D, 1 = H)					
2. Cost					
3. Amount of response (intensity)					
4. Choice to respond (frequency)					
5. Pen theft	1				
6. Felt obligation	.00	1			
7. Negative affect	-.04	-.19*	1		
8. Trust	-.05	.22**	-.60***	1	
9. Gender (0 = M, 1 = F)	.05	-.03	-.10	.11	1

Table 4. Regression estimations for trust and negative affect moderated mediation analyses (intensity), Experiment 2.

	<u>Dependent Variable</u>		
	Trust	Negative affect	Amount of Response
Partner's behavior			
<i>Honesty</i>	.71***	-.48***	.38***
<i>Deception (Ref)</i>			
Negative affect			.24 ⁺
Trust			-.07
Behavior x Trust			.20 ⁺
Behavior x Negative affect			-.26*
Gender			
<i>Female</i>	.10 ⁺	-.09	.11 ⁺
<i>Male (Ref)</i>			
Cost to respond	.01	.21**	-.32***
Adjusted R²	.50	.29	.30
Degrees of freedom	5	5	7

Note: The entries in these columns are standardized betas from the regression.

⁺ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p < .001$

Table 5. Logistic regression estimations for trust and negative affect moderated mediation analyses (pen theft), Experiment 2.

Independent Variables	β coefficient (SE)	Wald χ^2	Odds ratio
Partner's behavior			
<i>Honesty</i>	1.15 (.62) ⁺	3.47	3.19
<i>Deception (Ref)</i>			
Negative affect	.71 (.50)	2.03	2.04
Trust	.13 (.37)	.13	1.13
Behavior x Trust	-.98 (.48)*	4.09	.38
Behavior x Negative affect	-1.05 (.58) ⁺	3.27	.35
Cost to respond	-1.05 (.50)*	4.38	.35
Likelihood ratio χ^2	169.56		
Degrees of freedom	7		

⁺ $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p < .001$

Table 6. Means, standard deviations, and correlations for independent and dependent variables, Experiment 2 reanalyzed (N = 106)

	Mean	<i>SD</i>	1	2	3	4
1. Partner's behavior (0 = D, 1 = H)	.50	.50	1			
2. Ethnicity (0 = W, 1 = E)	.29	.48	.02	1		
3. Cost	1.00	.84	.11	.01	1	
4. Amount of response (intensity)	1.69	1.60	.43***	.02	-.40***	1
5. Choice to respond (frequency)	.64	.48	.51***	.14	-.32**	.79***
6. Felt obligation	3.37	1.76	.22*	.02	-.26**	.43***
7. Negative affect	3.46	1.06	-.47***	.09	.21*	-.28**
8. Trust	4.33	1.36	.74***	-.05	-.11	.41***
9. Gender (0 = M, 1 = F)	.49	.50	.08	-.05	.05	.10

⁺ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 6 continued. Means, standard deviations, and correlations for independent and dependent variables, Experiment 2 reanalyzed (N = 106)

	5	6	7	8	9
1. Partner's behavior (0 = D, 1 = H)					
2. Ethnicity (0 = W, 1 = E)					
3. Cost					
4. Amount of response (intensity)					
5. Choice to respond (frequency)	1				
6. Felt obligation	.43***	1			
7. Negative affect	-.23*	-.13*	1		
8. Trust	.46***	.18*	-.60***	1	
9. Gender (0 = M, 1 = F)	.10	-.12	-.12	.07	1

FIGURES

Figure 1. Model outlining responses to deception versus honesty.

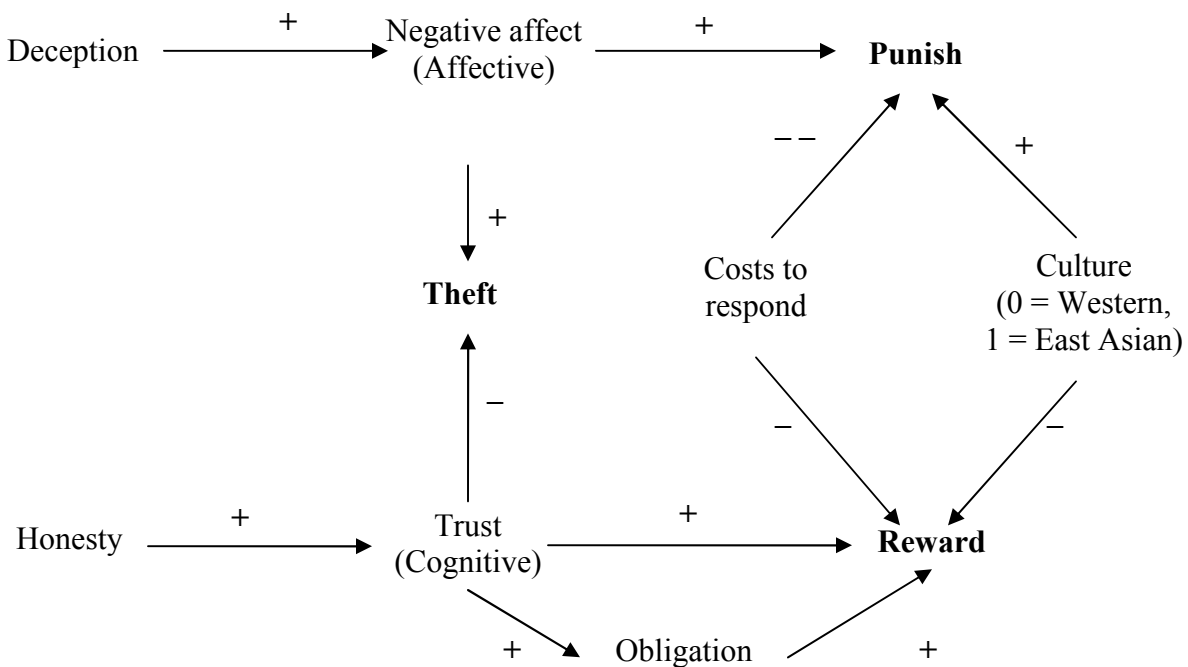
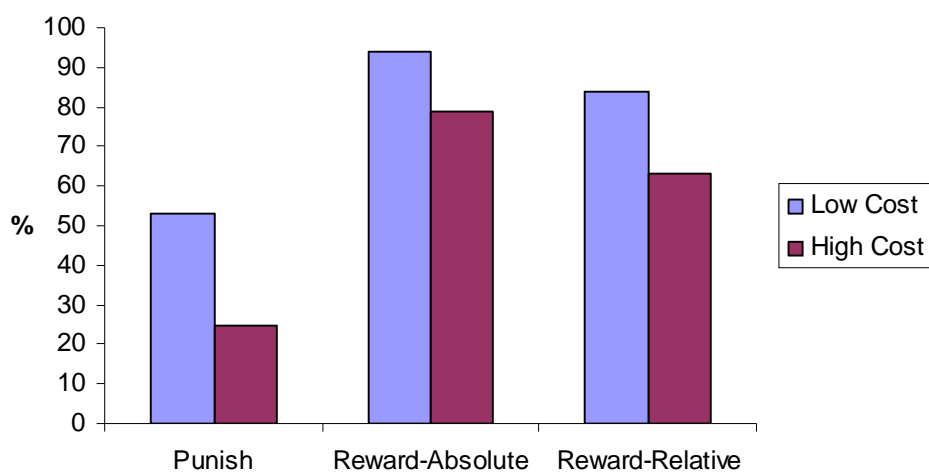
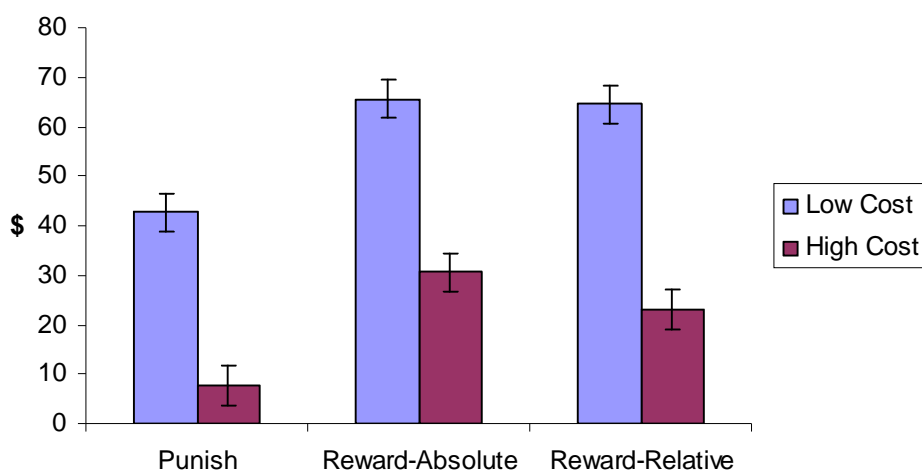


Figure 2. The effect of partner's behavior and cost on choice to respond (percentages and frequencies), Expt 1.



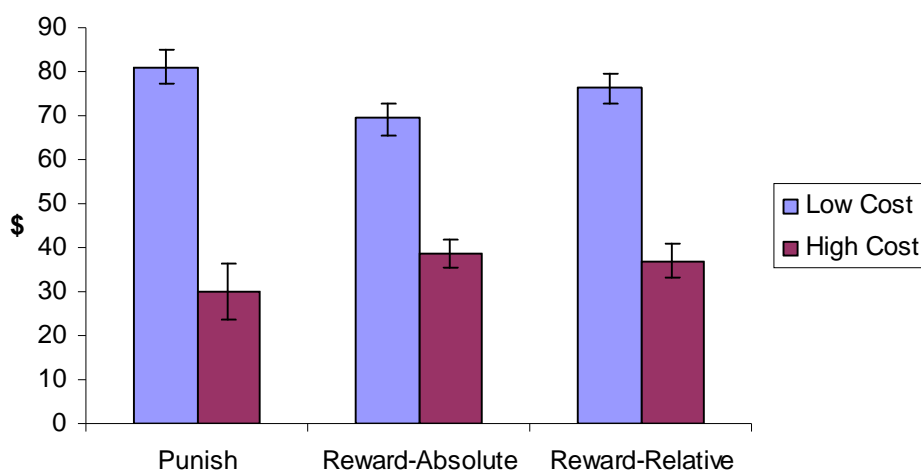
		Recipient's Action					
		Punish		Reward-Absolute		Reward-Relative	
Response Decision	Cost	Low	High	Low	High	Low	High
	Yes		10	4	17	15	16
No		9	12	1	4	3	6

Figure 3. The effect of partner's behavior and cost on amount of response, including zero responses (Means \pm SEM and frequencies per condition), Expt 1



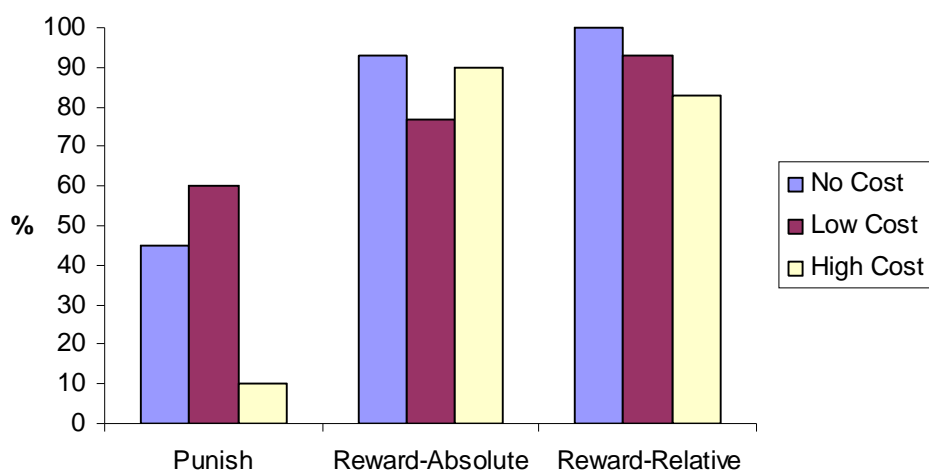
		Recipient's Action		
		Punish	Reward-Absolute	Reward-Relative
Cost	Low	19	18	19
	High	16	19	16

Figure 4. The effect of partner's behavior and cost on amount of response, excluding zero responses (Means \pm SEM and frequencies per condition), Expt 1



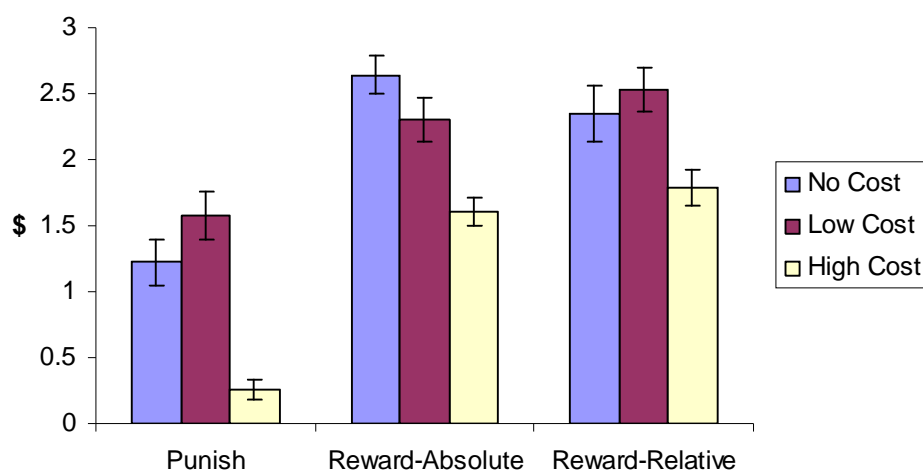
		Recipient's Action		
		Punish	Reward-Absolute	Reward-Relative
Cost	Low	10	17	16
	High	4	15	10

Figure 5. The effect of the other player's behavior and cost on choice to respond (percentages and frequencies), Expt 2.



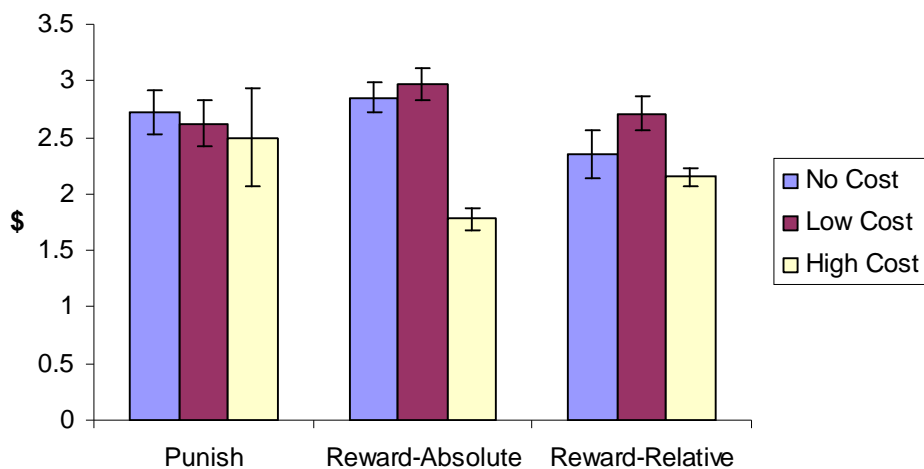
	Recipient's Action									
	Cost	Punish			Reward-Absolute			Reward-Relative		
		No	Low	High	No	Low	High	No	Low	High
Response Decision	Yes	9	12	3	26	17	18	10	14	10
No	11	8	26	2	5	2	0	1	2	

Figure 6. The effect of the other player's behavior and cost on intensity of response, including zero responses (Means \pm SEM and frequencies per condition), Expt 2.



		Recipient's Action		
		Punish	Reward-Absolute	Reward-Relative
Cost	None	20	28	10
	Low	20	22	15
	High	29	20	12

Figure 7. The effect of the other player's behavior and cost on intensity of response, excluding zero responses (Means \pm SEM and frequencies per condition), Expt 2.



		Recipient's Action		
		Punish	Reward-Absolute	Reward-Relative
Cost	None	9	26	10
	Low	12	17	14
	High	3	18	10

Figure 8. Trust and negative affect moderated mediation analyses (intensity), Expt 2.

Figure 8a.

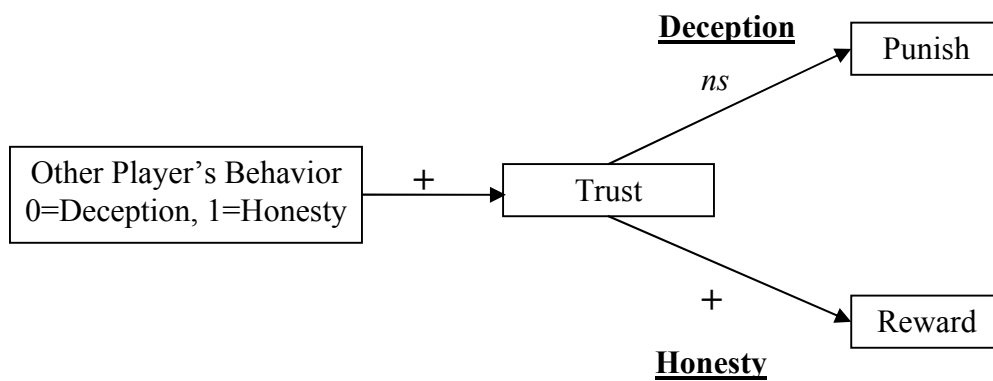


Figure 8b.

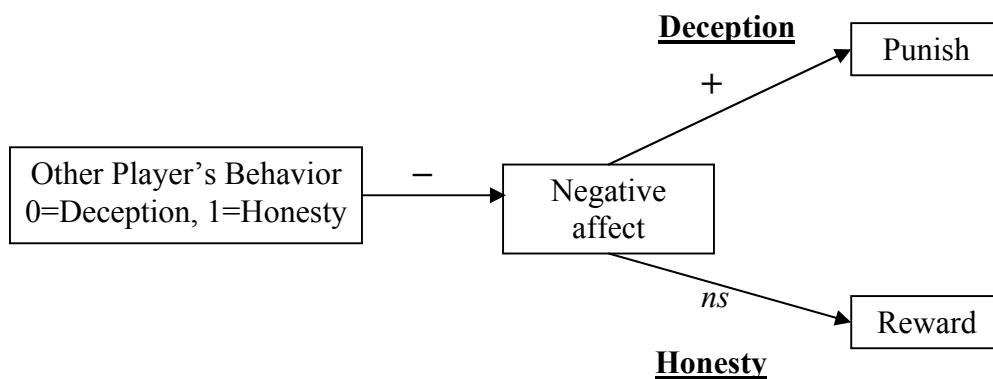


Figure 9. The effect of trust on intensity, Expt 2.

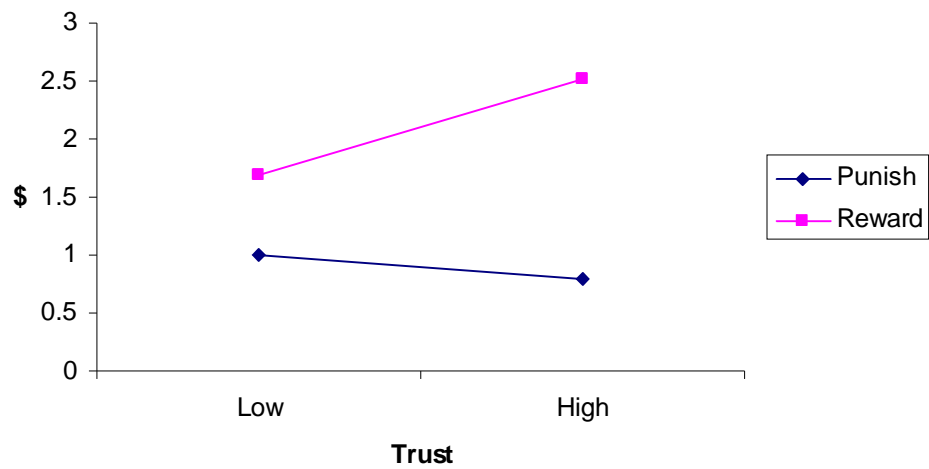


Figure 10. The effect of negative affect on intensity, Expt 2.



Figure 11. Felt obligation mediation analysis: Felt obligation as the mediator between trust and reward, Expt 2.

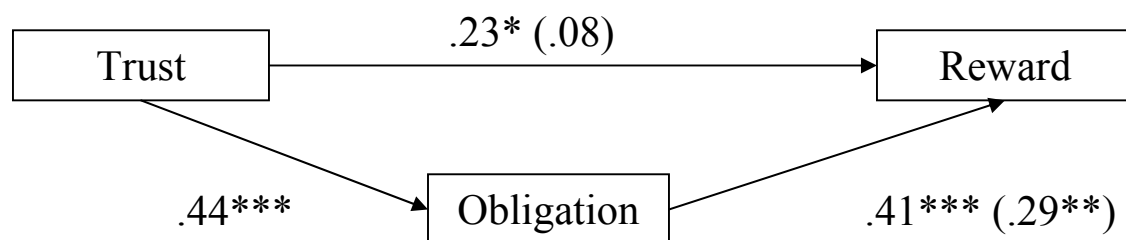
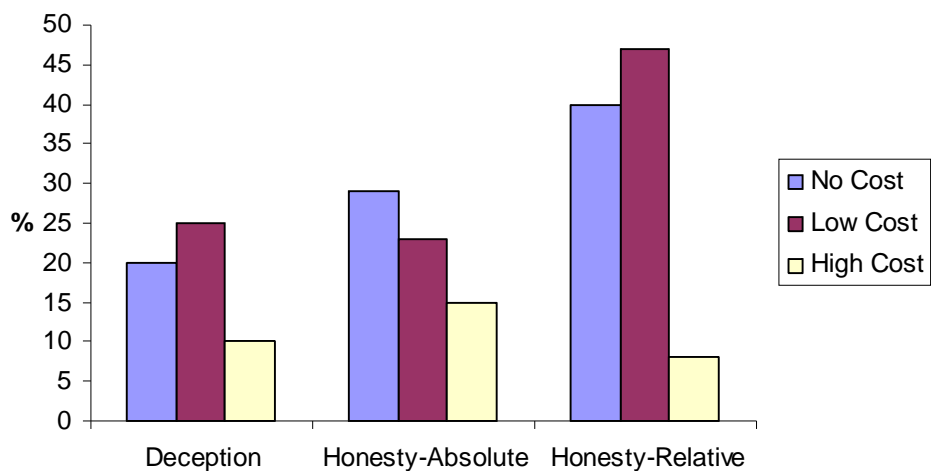


Figure 12. The effect of the other player's behavior and cost on pen theft, Expt 2.



		Other Player's Action								
		Deception			Honesty-Absolute			Honesty-Relative		
Took pen	Cost	No	Low	High	No	Low	High	No	Low	High
	No		16	15	26	20	17	17	6	8
Yes		4	5	3	8	5	3	4	7	1

Figure 13. Trust and negative affect moderated mediation analysis (pen theft), Expt 2.

Figure 13a.

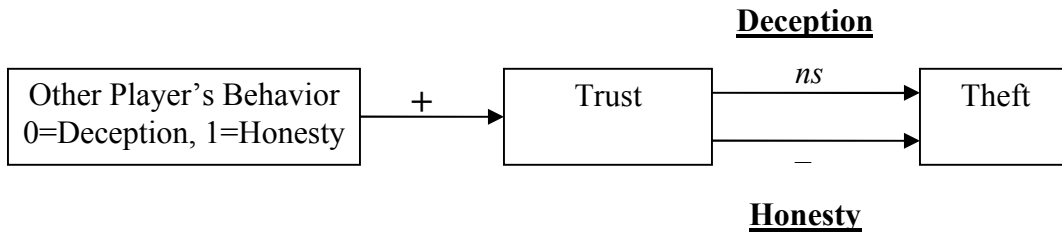
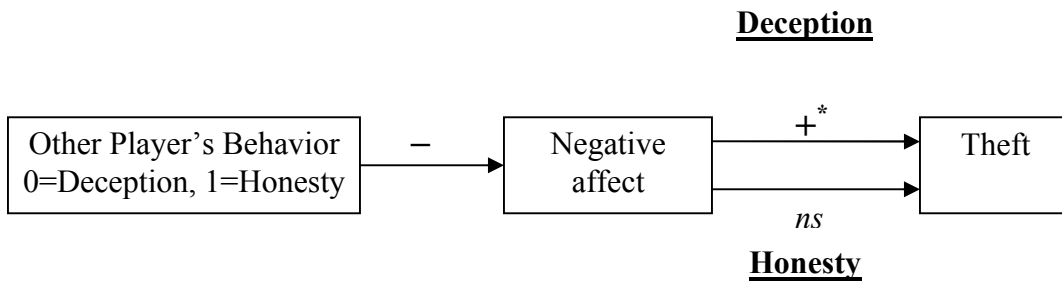
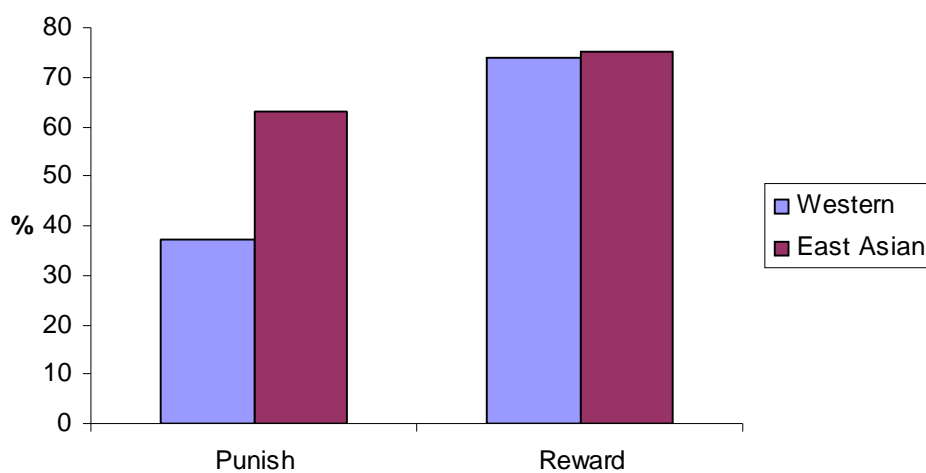


Figure 13b.



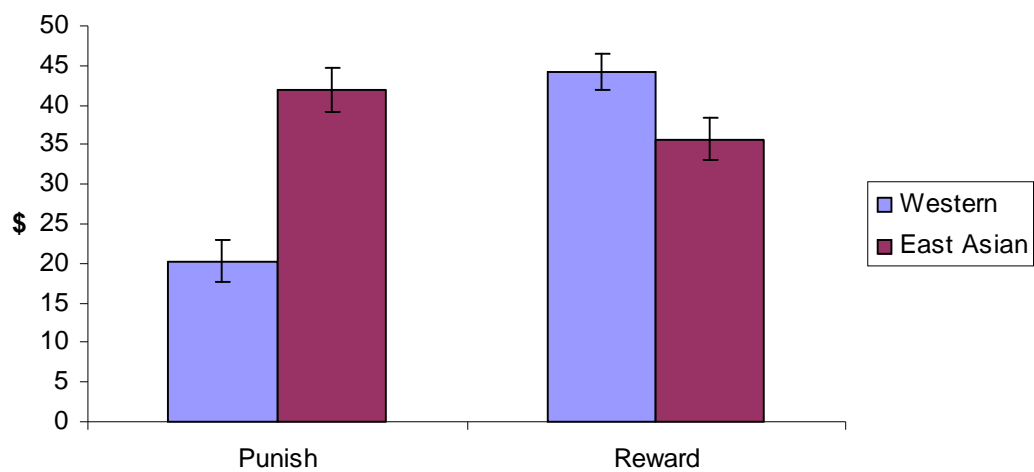
*Suggestive, but not statistically significant

Figure 14. The effect of partner's behavior and culture on choice to respond (percentages and frequencies), Expt 3.



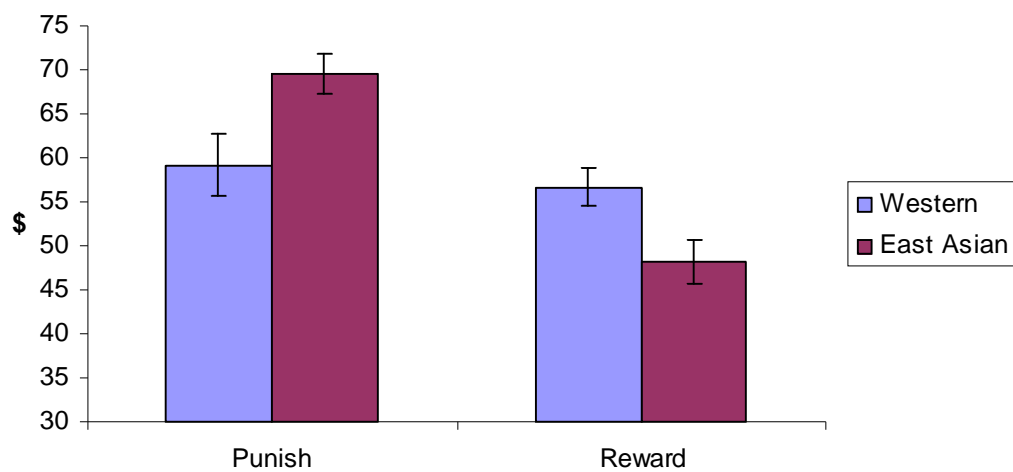
		Partner's Behavior			
		Deception		Honesty	
Response Decision	Culture	Western	East Asian	Western	East Asian
	Yes		13	30	49
No		22	18	17	7

Figure 15. The effect of partner's behavior and culture on amount of response, including zero responses (Means \pm SEM and frequencies per condition), Expt 3.



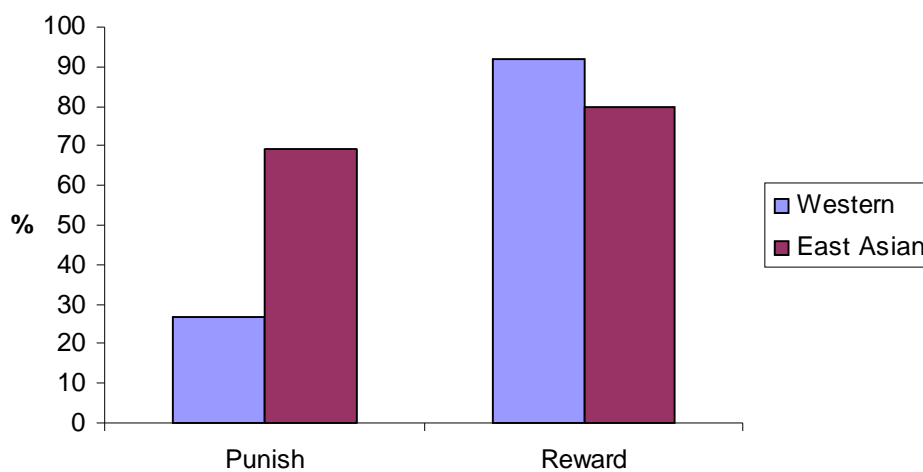
		Recipient's Action	
		Punish	Reward-Absolute
Culture	Western	35	66
	East Asian	48	28

Figure 16. The effect of partner's behavior and culture on amount of response, excluding zero responses (Means \pm SEM and frequencies per condition), Expt 3.



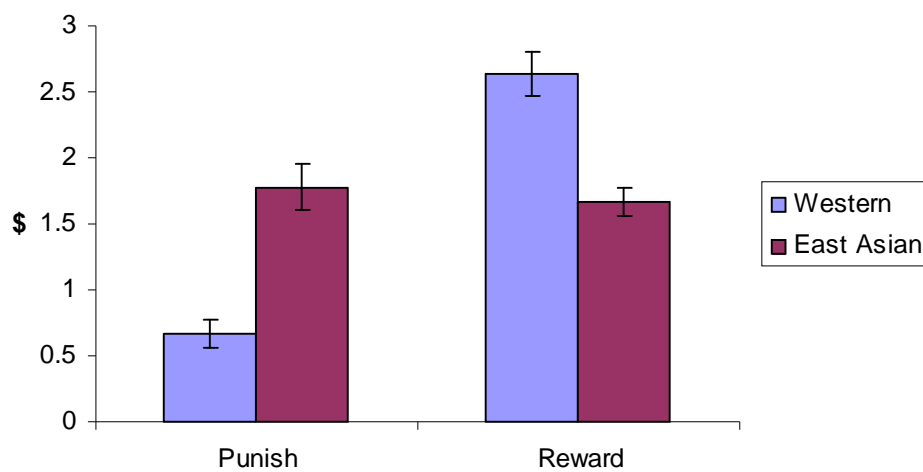
		Recipient's Action	
		Punish	Reward-Absolute
Culture	Western	13	49
	East Asian	30	21

Figure 17. The effect of partner's behavior and culture on frequency of response (percentages and frequencies), Expt 2 reanalyzed.



		Partner's Behavior			
		Deception		Honesty	
Response Decision	Culture	Western	East Asian	Western	East Asian
	Yes		10	11	35
No		27	5	3	3

Figure 18. The effect of partner's behavior and culture on intensity of response, including zero responses (Means \pm SEM and frequencies per condition), Expt 2 reanalyzed.



		Recipient's Action	
		Punish	Reward-Absolute
Culture	Western	37	38
	East Asian	16	15

Figure 19. The effect of partner's behavior and culture on intensity of response, excluding zero responses (Means \pm SEM and frequencies per condition), Expt 2 reanalyzed.

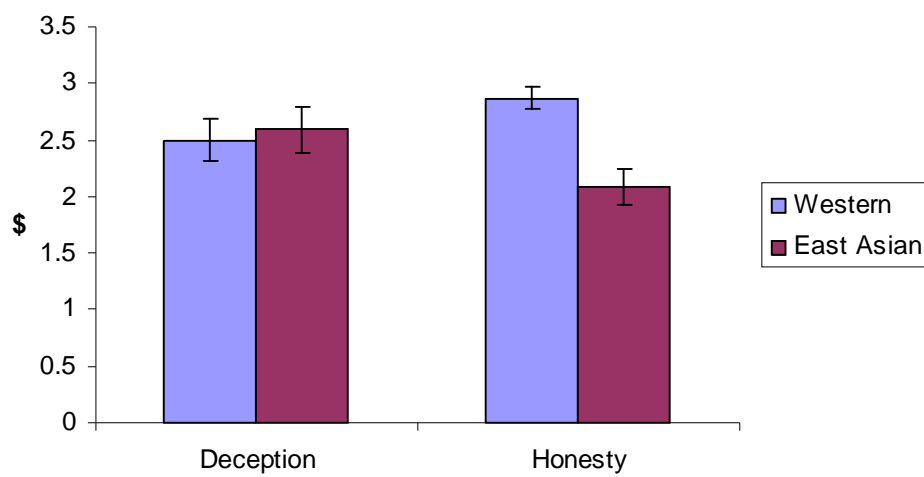


Figure 20. The effect of partner's behavior and culture on trust, Expt 2 reanalyzed.

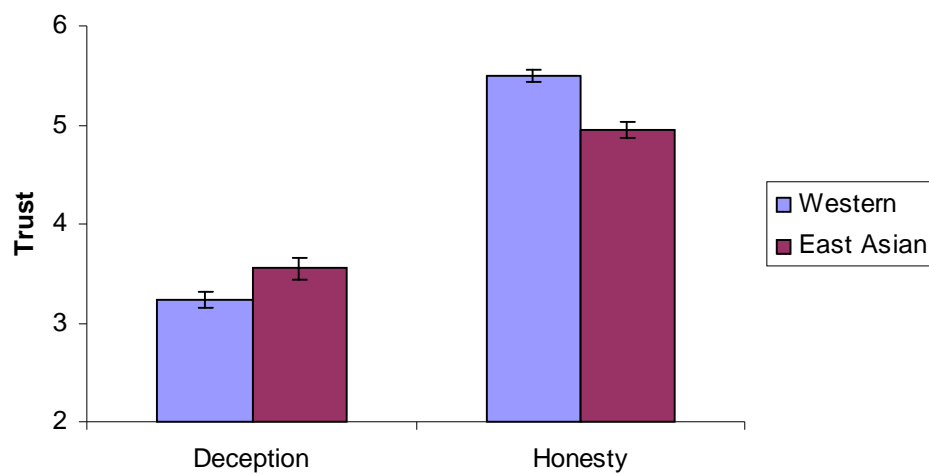


Figure 21. The effect of partner's behavior and culture on unhappiness, Expt 2 reanalyzed.

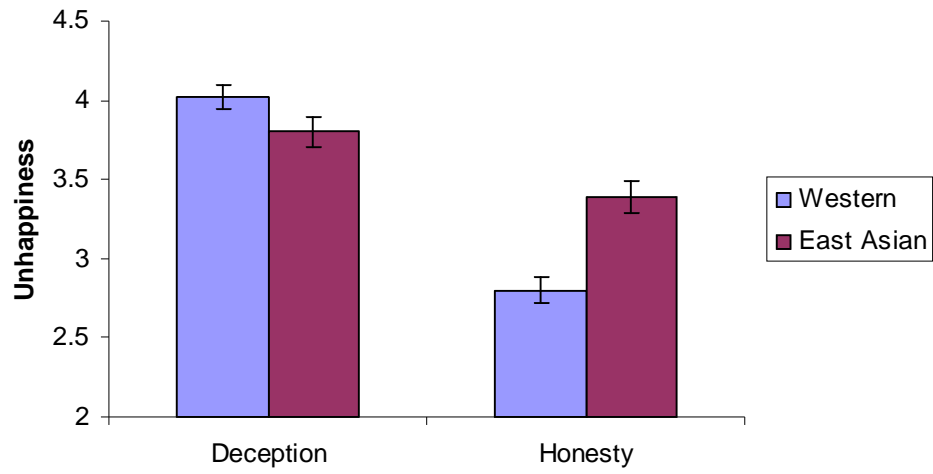
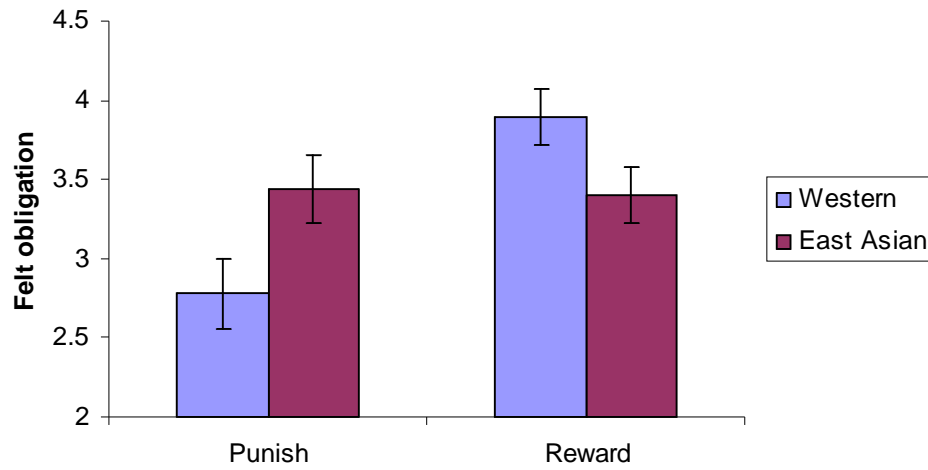


Figure 22. The effect of partner's behavior and culture on felt obligation, Expt 2 reanalyzed.



APPENDICES

Appendix 1: Scenarios Manipulating Pat's Behavior and Cost to Respond

Deception/Low Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was dishonest about some key information regarding the deal. As a result, you only received \$100. You would have received 50% more if Pat had given you honest information.

You have a one-time opportunity to punish Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to taking money from Pat at a 1:10 ratio. In other words, for every 10 cents you spend, you punish Pat 1 dollar. You can punish up to \$100.

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to subtract money from your bank. You will not receive any money taken from Pat.

1) How do you want to respond to Pat?

- Punish Pat \$100 (at a cost of \$10)
- Punish Pat \$90 (at a cost of \$9)
- Punish Pat \$80 (at a cost of \$8)
- Punish Pat \$70 (at a cost of \$7)
- Punish Pat \$60 (at a cost of \$6)
- Punish Pat \$50 (at a cost of \$5)
- Punish Pat \$40 (at a cost of \$4)
- Punish Pat \$30 (at a cost of \$3)
- Punish Pat \$20 (at a cost of \$2)
- Punish Pat \$10 (at a cost of \$1)
- Do not punish Pat

Appendix 1: Scenarios Manipulating Pat's Behavior and Cost to Respond (continued)

Honest-Relative/Low Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was honest about some key information regarding the deal. As a result, you received \$100. You would have received 25% less if Pat had given you dishonest information.

You have a one-time opportunity to reward Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to giving money to Pat at a 1:10 ratio. In other words, for every 10 cents you spend, you reward Pat 1 dollar. You can reward up to \$100.

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to add money to your bank.

1) How do you want to respond to Pat?

- Reward Pat \$100 (at a cost of \$10)
- Reward Pat \$90 (at a cost of \$9)
- Reward Pat \$80 (at a cost of \$8)
- Reward Pat \$70 (at a cost of \$7)
- Reward Pat \$60 (at a cost of \$6)
- Reward Pat \$50 (at a cost of \$5)
- Reward Pat \$40 (at a cost of \$4)
- Reward Pat \$30 (at a cost of \$3)
- Reward Pat \$20 (at a cost of \$2)
- Reward Pat \$10 (at a cost of \$1)
- Do not reward Pat

Appendix 1: Scenarios Manipulating Pat's Behavior and Cost to Respond (continued)

Honest-Absolute/Low Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was honest about some key information regarding the deal. As a result, you received \$100. You would have received 50% less if Pat had given you dishonest information.

You have a one-time opportunity to reward Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to giving money to Pat at a 1:10 ratio. In other words, for every 10 cents you spend, you reward Pat 1 dollar. You can reward up to \$100.

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to add money to your bank.

1) How do you want to respond to Pat?

- Reward Pat \$100 (at a cost of \$10)
- Reward Pat \$90 (at a cost of \$9)
- Reward Pat \$80 (at a cost of \$8)
- Reward Pat \$70 (at a cost of \$7)
- Reward Pat \$60 (at a cost of \$6)
- Reward Pat \$50 (at a cost of \$5)
- Reward Pat \$40 (at a cost of \$4)
- Reward Pat \$30 (at a cost of \$3)
- Reward Pat \$20 (at a cost of \$2)
- Reward Pat \$10 (at a cost of \$1)
- Do not reward Pat

Appendix 1: Scenarios Manipulating Pat's Behavior and Cost to Respond (continued)

Deception/High Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was dishonest about some key information regarding the deal. As a result, you only received \$100. You would have received 50% more if Pat had given you honest information.

You have a one-time opportunity to punish Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to taking money from Pat at a 1:1 ratio. In other words, for every 1 dollar you spend, you punish Pat 1 dollar. You can punish up to \$100.

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to subtract money from your bank. You will not receive any money taken from Pat.

1) How do you want to respond to Pat?

- Punish Pat \$100 (at a cost of \$100)
- Punish Pat \$90 (at a cost of \$90)
- Punish Pat \$80 (at a cost of \$80)
- Punish Pat \$70 (at a cost of \$70)
- Punish Pat \$60 (at a cost of \$60)
- Punish Pat \$50 (at a cost of \$50)
- Punish Pat \$40 (at a cost of \$40)
- Punish Pat \$30 (at a cost of \$30)
- Punish Pat \$20 (at a cost of \$20)
- Punish Pat \$10 (at a cost of \$10)
- Do not punish Pat

Appendix 1: Scenarios Manipulating Pat's Behavior and Cost to Respond (continued)

Honest-Relative/High Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was honest about some key information regarding the deal. As a result, you received \$100. You would have received 25% less if Pat had given you dishonest information.

You have a one-time opportunity to reward Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to giving money to Pat at a 1:1 ratio. In other words, for every 1 dollar you spend, you reward Pat 1 dollar. You can reward up to \$100.

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to add money to your bank.

1) How do you want to respond to Pat?

- Reward Pat \$100 (at a cost of \$100)
- Reward Pat \$90 (at a cost of \$90)
- Reward Pat \$80 (at a cost of \$80)
- Reward Pat \$70 (at a cost of \$70)
- Reward Pat \$60 (at a cost of \$60)
- Reward Pat \$50 (at a cost of \$50)
- Reward Pat \$40 (at a cost of \$40)
- Reward Pat \$30 (at a cost of \$30)
- Reward Pat \$20 (at a cost of \$20)
- Reward Pat \$10 (at a cost of \$10)
- Do not reward Pat

Appendix 1: Scenarios Manipulating Pat's Behavior and Cost to Respond (continued)

Honest-Absolute/High Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was honest about some key information regarding the deal. As a result, you received \$100. You would have received 50% less if Pat had given you dishonest information.

You have a one-time opportunity to reward Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to giving money to Pat at a 1:1 ratio. In other words, for every 1 dollar you spend, you reward Pat 1 dollar. You can reward up to \$100.

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to add money to your bank.

1) How do you want to respond to Pat?

- Reward Pat \$100 (at a cost of \$100)
- Reward Pat \$90 (at a cost of \$90)
- Reward Pat \$80 (at a cost of \$80)
- Reward Pat \$70 (at a cost of \$70)
- Reward Pat \$60 (at a cost of \$60)
- Reward Pat \$50 (at a cost of \$50)
- Reward Pat \$40 (at a cost of \$40)
- Reward Pat \$30 (at a cost of \$30)
- Reward Pat \$20 (at a cost of \$20)
- Reward Pat \$10 (at a cost of \$10)
- Do not reward Pat

Appendix 2: Stage 1 Instructions

In Stage 1 and Stage 2, you will be matched with another person from this session. You will work with the same person in both stages. The other person will be PLAYER A and you will be PLAYER B. Neither of you will ever know the identity of the other. Two possible monetary payments are available to you (PLAYER B) and the other person (PLAYER A) in this stage. These payments for you and the other person depend on the options you choose. You will be given two options to choose from, but ONLY PLAYER A will know the actual amounts of each option.

To illustrate the procedure, imagine that payments in Stage 1 are given in a fictitious currency called MAXS. There are three parts to Stage 1:

- 1) PLAYER A (the other person) will be given information about two options. An example of the options might be the following:
OPTION A is that 'Player A receives 10 MAXS and Player B receives 5 MAXS'.
OPTION B is that 'Player A receives 8 MAXS and Player B receives 12 MAXS'.
- 2) PLAYER A can then send either of two possible messages:
Message 1 (Not the truth): "Option A will earn you more money than option B"
OR
Message 2 (The truth): "Option B will earn you more money than option A"
- 3) PLAYER B (you) will choose between Option A and Option B. But remember, you will not know the amounts in each option. The ONLY INFORMATION you will have before making a choice is the message that Player A sends.

Please wait for the experimenter to give you Player A's message.

Appendix 3: Stage 2 Instructions and Choice

Punish/No Cost Manipulation

Stage 2

You have a one-time opportunity to punish Player A. Punishments can be at most 4 dollars.

After your choice, you WILL NOT interact with Player A again. Player A WILL NOT have the option to subtract money from your bank. You will not receive any money taken from Player A.

What is your choice?

- Punish Player A \$4.00
- Punish Player A \$3.50
- Punish Player A \$3.00
- Punish Player A \$2.50
- Punish Player A \$2.00
- Punish Player A \$1.50
- Punish Player A \$1.00
- Punish Player A \$0.50
- Do not punish Player A

Appendix 3: Stage 2 Instructions and Choice (continued)

Reward/No Cost Manipulation

Stage 2

You have a one-time opportunity to reward Player A. Rewards can be at most 4 dollars.

After your choice, you WILL NOT interact with Player A again. Player A WILL NOT have the option to add money to your bank.

What is your choice?

- Reward Player A \$4.00
- Reward Player A \$3.50
- Reward Player A \$3.00
- Reward Player A \$2.50
- Reward Player A \$2.00
- Reward Player A \$1.50
- Reward Player A \$1.00
- Reward Player A \$0.50
- Do not reward Player A

Appendix 3: Stage 2 Instructions and Choice (continued)

Punish/Low Cost Manipulation

Stage 2

You have a one-time opportunity to punish Player A, your choice will require you to spend your own money from Stage 1. For every 10 cents you spend, you can punish Player A 1 dollar. Punishments can be at most 4 dollars.

After your choice, you WILL NOT interact with Player A again. Player A WILL NOT have the option to subtract money from your bank. You will not receive any money taken from Player A.

What is your choice?

- Punish Player A \$4.00 (at a cost to you of \$.40)
- Punish Player A \$3.50 (at a cost to you of \$.35)
- Punish Player A \$3.00 (at a cost to you of \$.30)
- Punish Player A \$2.50 (at a cost to you of \$.25)
- Punish Player A \$2.00 (at a cost to you of \$.20)
- Punish Player A \$1.50 (at a cost to you of \$.15)
- Punish Player A \$1.00 (at a cost to you of \$.10)
- Punish Player A \$0.50 (at a cost to you of \$.05)
- Do not punish Player A

Appendix 3: Stage 2 Instructions and Choice (continued)

Reward/Low Cost Manipulation

Stage 2

You have a one-time opportunity to reward Player A, your choice will require you to spend your own money from Stage 1. For every 10 cents you spend, you can reward Player A 1 dollar. Rewards can be at most 4 dollars.

After your choice, you WILL NOT interact with Player A again. Player A WILL NOT have the option to add money to your bank.

What is your choice?

- Reward Player A \$4.00 (at a cost to you of \$.40)
- Reward Player A \$3.50 (at a cost to you of \$.35)
- Reward Player A \$3.00 (at a cost to you of \$.30)
- Reward Player A \$2.50 (at a cost to you of \$.25)
- Reward Player A \$2.00 (at a cost to you of \$.20)
- Reward Player A \$1.50 (at a cost to you of \$.15)
- Reward Player A \$1.00 (at a cost to you of \$.10)
- Reward Player A \$0.50 (at a cost to you of \$.05)
- Do not reward Player A

Appendix 3: Stage 2 Instructions and Choice (continued)

Punish/High Cost Manipulation

Stage 2

You have a one-time opportunity to punish Player A, your choice will require you to spend your own money from Stage 1. For every 1 dollar you spend, you can punish Player A 1 dollar. Punishments can be at most 4 dollars.

After your choice, you WILL NOT interact with Player A again. Player A WILL NOT have the option to subtract money from your bank. You will not receive any money taken from Player A.

What is your choice?

- Punish Player A \$4.00 (at a cost to you of \$4.00)
- Punish Player A \$3.50 (at a cost to you of \$3.50)
- Punish Player A \$3.00 (at a cost to you of \$3.00)
- Punish Player A \$2.50 (at a cost to you of \$2.50)
- Punish Player A \$2.00 (at a cost to you of \$2.00)
- Punish Player A \$1.50 (at a cost to you of \$1.50)
- Punish Player A \$1.00 (at a cost to you of \$1.00)
- Punish Player A \$0.50 (at a cost to you of \$0.50)
- Do not punish Player A

Appendix 3: Stage 2 Instructions and Choice (continued)

Reward/High Cost Manipulation

Stage 2

You have a one-time opportunity to reward Player A, your choice will require you to spend your own money from Stage 1. For every 1 dollar you spend, you can reward Player A 1 dollar. Rewards can be at most 4 dollars.

After your choice, you WILL NOT interact with Player A again. Player A WILL NOT have the option to add money to your bank.

What is your choice?

- Reward Player A \$4.00 (at a cost to you of \$4.00)
- Reward Player A \$3.50 (at a cost to you of \$3.50)
- Reward Player A \$3.00 (at a cost to you of \$3.00)
- Reward Player A \$2.50 (at a cost to you of \$2.50)
- Reward Player A \$2.00 (at a cost to you of \$2.00)
- Reward Player A \$1.50 (at a cost to you of \$1.50)
- Reward Player A \$1.00 (at a cost to you of \$1.00)
- Reward Player A \$0.50 (at a cost to you of \$0.50)
- Do not reward Player A

Appendix 4: Dependent psychological measures

Negative affect

1. After you found out the information about the other player's choice in Stage 1, to what degree did you feel the following emotion? HAPPY

1	2	3	4	5	6	7
not at all	slightly	somewhat	moderately	very	extremely	completely

2. After you found out the information about the other player's choice in Stage 1, to what degree did you feel the following emotion? UNHAPPY

1	2	3	4	5	6	7
not at all	slightly	somewhat	moderately	very	extremely	completely

3. After you found out the information about the other player's choice in Stage 1, did you feel:

1	2	3	4	5	6	7
extremely disgruntled	moderately disgruntled	slightly disgruntled	neither disgruntled nor pleased	slightly pleased	moderately pleased	extremely pleased

4. After you found out the information about the other player's choice in Stage 1, did you feel:

1	2	3	4	5	6	7
extremely happy	moderately happy	slightly happy	neither happy nor unhappy	slightly unhappy	moderately unhappy	extremely unhappy

5. When making the decision in Stage 2, I felt:

extremely pleased	moderately pleased	slightly pleased	neither disgruntled nor pleased	slightly disgruntled	moderately disgruntled	extremely disgruntled
----------------------	-----------------------	---------------------	---------------------------------------	-------------------------	---------------------------	--------------------------

6. When making the decision in Stage 2, I felt:

1	2	3	4	5	6	7
extremely unhappy	moderately unhappy	slightly unhappy	neither happy nor unhappy	slightly happy	moderately happy	extremely happy

Appendix 4: Dependent psychological measures

Measures of Trust

1. I consider the other player to be _____.

1	2	3	4	5	6	7
extremely unfair	moderately unfair	slightly unfair	neither fair nor unfair	slightly fair	moderately fair	extremely fair

2. I consider the other player to be _____.

1	2	3	4	5	6	7
extremely unkind	moderately unkind	slightly unkind	neither kind nor unkind	slightly kind	moderately kind	extremely kind

3. I consider the other player to be _____.

1	2	3	4	5	6	7
extremely cooperative	moderately cooperative	slightly cooperative	neither cooperative nor uncooperative	slightly uncooperative	moderately uncooperative	extremely uncooperative

4. I consider the other player to be _____.

1	2	3	4	5	6	7
extremely trustworthy	moderately trustworthy	slightly trustworthy	neither trustworthy nor untrustworthy	slightly untrustworthy	moderately untrustworthy	extremely untrustworthy

Measures of Felt Obligation

When making the decision in Stage 2, I felt:

1	2	3	4	5	6	7
completely obligated to react	extremely obligated to react	very obligated to react	moderately happy nor unhappy	somewhat happy	slightly happy	not at all happy

Appendix 5: Scenarios Manipulating Pat's Behavior and Cost to Respond, Taiwanese version
Dishonest/Low Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

請回答如下的問題。你的答案將是完全保密的。

Imagine the following scenario: You and Pat recently completed a deal and you have just found out that Pat was dishonest about some key information regarding the deal. As a result, you only received \$100 (approximately NT\$ 3300). You would have received 50% more if Pat had given you honest information. 想像如下一個場景。你和派特（人名）完成了一項商業交易，你卻發現對於這個協定的一些重要的訊息，派特並不誠實。結果你僅僅獲得了100美圓（大約3300臺幣）。如果派特提供給你真實的訊息，你將多獲得50%。

You have a one-time opportunity to punish Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to giving taking money from Pat at a 1:10 ratio. In other words, for every 10 cents you spend, you punish Pat 1 dollar. You can punish up to \$100.

你有一次機會可以懲罰派特，但是需要你動用你自己的錢。你可以選擇用一種方式，相當於按照1：10的比率從派特那裏拿走錢。換句話說，你每花10美分（10 cents），你可以懲罰派特1美圓。你可以最多懲罰派特100美圓。

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to subtract money from your bank. You will not receive any money taken from Pat.

在你選擇之後，你不會再和派特有任何的關係。派特也不能選擇向你的銀行裏增加或者提取錢。你不會得到任何派特已經拿走的錢。

1) How do you want to respond to Pat? 你想如何回應派特？

- Punish Pat \$100 (at a cost of \$10) 懲罰派特100美圓 (花費你自己10美圓)。
- Punish Pat \$90 (at a cost of \$9) 懲罰派特90美圓 (花費你自己9美圓)。
- Punish Pat \$80 (at a cost of \$8) 懲罰派特80美圓 (花費你自己8美圓)。
- Punish Pat \$70 (at a cost of \$7) 懲罰派特70美圓 (花費你自己7美圓)。
- Punish Pat \$60 (at a cost of \$6) 懲罰派特60美圓 (花費你自己6美圓)。
- Punish Pat \$50 (at a cost of \$5) 懲罰派特50美圓 (花費你自己5美圓)。
- Punish Pat \$40 (at a cost of \$4) 懲罰派特40美圓 (花費你自己4美圓)。
- Punish Pat \$30 (at a cost of \$3) 懲罰派特30美圓 (花費你自己3美圓)。
- Punish Pat \$20 (at a cost of \$2) 懲罰派特20美圓 (花費你自己2美圓)。
- Punish Pat \$10 (at a cost of \$2) 懲罰派特10美圓 (花費你自己1美圓)。
- Do not punish Pat 不懲罰派特。

Appendix 5: Scenarios Manipulating Pat's Behavior and Cost to Respond, Taiwanese version

(Continued) Honest/Low Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was honest about some key information regarding the deal. As a result, you received \$100 (approximately NT\$ 3300). You would have received 50% less if Pat had given you dishonest information.

想像如下一個場景。你和派特（人名）完成了一項商業交易，你卻發現對於這個協定的一些重要的訊息，派特是誠實的。結果你獲得了100美圓（大約3300臺幣）。如果派特提供給你不誠實的訊息，你將少獲得50%。

You have a one-time opportunity to reward Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to giving money to Pat at a 1:10 ratio. In other words, for every 10 cents you spend, you reward Pat 1 dollar. You can reward up to \$100. 你有一次機會可以獎勵派特，但是需要你動用你自己的錢。你可以選擇用一種方式，相當於按照1：10的比率給于派特錢。換句話說，你每花10美分（10 cents），你可以獎勵派特1美圓。你可以最多獎勵派特100美圓。

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to add money to your bank.

在你選擇之後，你不會再和派特有任何的關係。派特也不能選擇向你的銀行裏增加錢。

1) How do you want to respond to Pat? 你想如何回應派特?

- Reward Pat \$100 (at a cost of \$10) 獎勵派特100美圓（花費你自己10美圓）
- Reward Pat \$90 (at a cost of \$9) 獎勵派特90美圓（花費你自己9美圓）
- Reward Pat \$80 (at a cost of \$8) 獎勵派特80美圓（花費你自己8美圓）
- Reward Pat \$70 (at a cost of \$7) 獎勵派特70美圓（花費你自己7美圓）
- Reward Pat \$60 (at a cost of \$6) 獎勵派特60美圓（花費你自己6美圓）
- Reward Pat \$50 (at a cost of \$5) 獎勵派特50美圓（花費你自己5美圓）
- Reward Pat \$40 (at a cost of \$4) 獎勵派特40美圓（花費你自己4美圓）
- Reward Pat \$30 (at a cost of \$3) 獎勵派特30美圓（花費你自己3美圓）
- Reward Pat \$20 (at a cost of \$2) 獎勵派特20美圓（花費你自己2美圓）
- Reward Pat \$10 (at a cost of \$1) 獎勵派特10美圓（花費你自己1美圓）
- Do not reward Pat 不獎勵派特

Appendix 5: Scenarios Manipulating Pat's Behavior and Cost to Respond, Taiwanese version
(continued) Dishonest/High Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was dishonest about some key information regarding the deal. As a result, you only received \$100 (approximately NT\$ 3300). You would have received 50% more if Pat had given you honest information.

想像如下一個場景。你和派特（人名）完成了一項商業交易，你卻發現對於這個協定的一些重要的訊息，派特並不誠實。結果你僅僅獲得了100美圓（大約3300臺幣）。如果派特提供給你真實的訊息，你將多獲得50%，。

You have a one-time opportunity to punish Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to taking money from Pat at a 1:1 ratio. In other words, for every 1 dollar you spend, you punish Pat 1 dollar. You can punish up to \$100.

你有一次機會可以懲罰派特，但是需要你動用你自己的錢。你可以選擇用一種方式，相當於按照1：1的比率從派特那裏拿走錢。換句話說，你每花1美分，你可以獎勵或者懲罰派特1美圓。你可以最多懲罰派特100美圓。

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to subtract money from your bank. You will not receive any money taken from Pat.

在你選擇之後，你不會再和派特有任何的關係。派特也不能選擇向你的銀行裏增加或者提取錢。你不會得到任何派特已經拿走的錢。

1) How do you want to respond to Pat? 你想如何回應派特?

- Punish Pat \$100 (at a cost of \$100) 懲罰派特100美圓 (花費你自己100美圓)。
- Punish Pat \$90 (at a cost of \$90) 懲罰派特90美圓 (花費你自己90美圓)。
- Punish Pat \$80 (at a cost of \$80) 懲罰派特80美圓 (花費你自己80美圓)。
- Punish Pat \$70 (at a cost of \$70) 懲罰派特70美圓 (花費你自己70美圓)。
- Punish Pat \$60 (at a cost of \$60) 懲罰派特60美圓 (花費你自己60美圓)。
- Punish Pat \$50 (at a cost of \$50) 懲罰派特50美圓 (花費你自己50美圓)。
- Punish Pat \$40 (at a cost of \$40) 懲罰派特40美圓 (花費你自己40美圓)。
- Punish Pat \$30 (at a cost of \$30) 懲罰派特30美圓 (花費你自己30美圓)。
- Punish Pat \$20 (at a cost of \$20) 懲罰派特20美圓 (花費你自己20美圓)。
- Punish Pat \$10 (at a cost of \$10) 懲罰派特10美圓 (花費你自己10美圓)。
- Do not punish Pat 不懲罰派特。

Appendix 5: Scenarios Manipulating Pat's Behavior and Cost to Respond, Taiwanese version
(continued) Honest/High Cost Manipulation

Please answer the following questions. Your answers are completely confidential.

Imagine the following scenario: You and Pat recently completed a business deal and you have just found out that Pat was honest about some key information regarding the deal. As a result, you received \$100 (approximately NT\$ 3300). You would have received 50% less if Pat had given you dishonest information.

想像如下一個場景。你的派特（人名）完成了一項商業交易，你卻發現對於這個協定的一些重要的訊息，派特是誠實的。結果你獲得了100美圓（大約3300臺幣）。如果派特提供給你不誠實的訊息，你將少獲得50%。

You have a one-time opportunity to reward or punish Pat, but responding will require you to spend your own money. You can choose to behave in a fashion that is equivalent to giving money to or taking money from Pat at a 1:1 ratio. In other words, for every 1 dollar you spend, you reward or punish Pat 1 dollar. You can reward or punish up to \$100.

你有一次機會可以獎勵派特，但是需要你動用你自己的錢。你可以選擇用一種方式，相當於按照1：1的比率給予派特錢。換句話說，你每花1美分，你可以獎勵派特1美圓。你可以最多獎勵派特100美圓。

After your choice, you WILL NOT interact with Pat again. Pat WILL NOT have the option to add/subtract money from your bank.

在你選擇之後，你不會再和派特有任何的關係。派特也不能選擇向你的銀行裏增加錢。

1) How do you want to respond to Pat? 你想如何回應派特?

- Reward Pat \$100 (at a cost of \$100) 獎勵派特100美圓（花費你自己100美圓）
- Reward Pat \$90 (at a cost of \$90) 獎勵派特90美圓（花費你自己90美圓）
- Reward Pat \$80 (at a cost of \$80) 獎勵派特80美圓（花費你自己80美圓）
- Reward Pat \$70 (at a cost of \$70) 獎勵派特70美圓（花費你自己70美圓）
- Reward Pat \$60 (at a cost of \$60) 獎勵派特60美圓（花費你自己60美圓）
- Reward Pat \$50 (at a cost of \$50) 獎勵派特50美圓（花費你自己50美圓）
- Reward Pat \$40 (at a cost of \$40) 獎勵派特40美圓（花費你自己40美圓）
- Reward Pat \$30 (at a cost of \$30) 獎勵派特30美圓（花費你自己30美圓）
- Reward Pat \$20 (at a cost of \$20) 獎勵派特20美圓（花費你自己20美圓）
- Reward Pat \$10 (at a cost of \$10) 獎勵派特10美圓（花費你自己10美圓）
- Do not reward Pat 不獎勵派特