Bad Drives Psychological Reactions, but Good Propels Behavior : Responses to Honesty and Deception

Cynthia S. Wang, Adam D. Galinsky and J. Keith Murnighan

Psychological Science 2009 20: 634
DOI: 10.1111/j.1467-9280.2009.02344.x

A JOURNAL OF THE ASSOCIATION FOR PSYCHOLOGICAL SCIENCE aps

Research Article

Bad Drives Psychological Reactions, but Good Propels Behavior

Responses to Honesty and Deception

Cynthia S. Wang,[1] Adam D. Galinsky,[2] and J. Keith Murnighan[2]

[1]National University of Singapore and [2]Northwestern University

ABSTRACT—*Research across disciplines suggests that bad is stronger than good and that individuals punish deception more than they reward honesty. However, methodological issues in previous research limit the latter conclusion. Three experiments resolved these issues and consistently found the opposite pattern: Individuals rewarded honesty more frequently and intensely than they punished deception. Experiment 2 extended these counterintuitive findings by revealing a divergence between evaluation and behavior: Evaluative reactions to deception were stronger than those to honesty, but behavioral intentions in response to honesty were stronger than those in response to deception. In addition, individuals wanted to avoid deceivers more than they wanted to approach honest actors. Experiment 3 found that punishment, but not reward, frequencies were sensitive to costs. Moderated-mediation tests revealed the role of different psychological mechanisms: Negative affect drove punishments, whereas perceived trustworthiness drove rewards. Overall, bad appears to be stronger than good in influencing psychological reactions, but good seems to be stronger than bad in influencing behavior.*

Societies, organizations, and leaders all face an enduring dilemma: how to effectively encourage moral behavior. Since Plato, a simple strategy has been used to achieve this goal: reward honesty and punish deception. Several disciplines, including the study of social psychology (Eisenberger, Lynch, Aselage, & Rohdieck, 2004; Gouldner, 1960), economics (Fehr & Gachter, 2000), and organizational behavior (Cropanzano & Mitchell, 2005), have investigated the effectiveness of this approach, and their accumulated results and theory are completely consistent. People punish deception more than they reward honesty because negative events psychologically outweigh positive events. Despite this widespread consensus, the results reported here contradict this conclusion. Three experiments consistently found that people rewarded honesty more than they punished deception, even when the choice of either action entailed real, substantive costs.

Address correspondence to Cynthia S. Wang, Department of Management and Organisation, NUS Business School, 1 Business Link, Singapore 117592, e-mail: cswang@nus.edu.sg.

BADS ARE STRONGER THAN GOODS: EVIDENCE FROM PSYCHOLOGY AND ECONOMICS

The psychology literature has repeatedly and broadly demonstrated that negatives are more powerful than positives (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001). Negative events influence emotions, cognitions, and behavior more than positive events (Rozin & Royzman, 2001). People process negative information more thoroughly than positive information (Dreben, Fiske, & Hastie, 1979), remember negative behaviors more accurately and vividly than positive behaviors (Fiske, 1980), and overemphasize negative data in impression formation (Falk & Fischbacher, 2006; Peeters & Czapinski, 1990).

Experimental economics has extended this research by documenting that people more readily punish deception than reward honesty (Abbink, Irlenbusch, & Renner, 2000; Brandts & Charness, 2003; Fehr & Gachter, 2000; Offerman, 2002), even in one-shot interactions with no possibility for future gain (Fehr & Gachter, 2000). The accumulated data are strikingly consistent: Bad trumps good, and, as a result, people punish deception more than they reward honesty.

Psychological Science RESEARCH, THEORY, & APPLICATION IN PSYCHOLOGY AND RELATED SCIENCES

Several methodological confounds, however, make it difficult to draw clear behavioral conclusions from this research. In these experiments, action is sequential: First movers engage in a behavior that either hurts or helps second movers, and these second movers can then respond by punishing or rewarding the first movers. Offerman (2002), for example, reported that second movers reacted more frequently to being intentionally harmed than to being helped; however, because initial actions led to different outcomes (first movers received 11 guilders for hurtful behavior but only 8 guilders for helpful behavior), it is easy to see why punishments were more frequent. In the study by Brandts and Charness (2003), second movers punished 56% of lying first movers (by decreasing their payoffs), but rewarded only 25% of truth tellers. However, for every unit second movers spent, they could punish by 10 units but reward by only 1 unit. Similarly, in Abbink et al. (2000), for every unit second movers spent, they could punish by 3 units but reward by only 1 unit. Because differential costs and unequal consequences could have encouraged more punishments than rewards, it is difficult to reach confident conclusions.

OVERVIEW OF THE PRESENT STUDY

We investigated responses to deception and honesty in three experiments that equalized the costs and effects of rewards and punishments. We predicted that people would reward honesty more frequently and with greater intensity than they would punish deception. Before describing our experiments, we offer both a theoretical reconciliation with the "bad is stronger than good" findings and an overview of our methodological improvements that overcome previous confounds.

Theoretical Reconciliation

At first blush, stronger reactions to honesty than to deception would seem to contradict the ubiquitous finding that "bad is stronger than good." Upon closer inspection, however, these findings may be complementary rather than contradictory. Because punishments tend to hurt more than rewards help (Baumeister et al., 2001), people may attempt to use rewards more frequently so that they will have an effect. In contrast, because bad affects people more intensely than good and everyone has experienced the pain of punishment, people realize that punishments have a stronger impact than rewards, so they tend to use punishments sparingly. Thus, good may influence direct behavioral responses more than bad because negatives produce stronger psychological reactions than positives.

Methodological Improvements

Given the confounds of past experimental designs, we took steps to equate costs and outcomes. If it costs less to punish than to reward, for example, people might choose to punish because it is cheaper. If initial outcomes differ (e.g., deception results in a $5 loss, but honesty results in a $3 gain), people might choose to punish more than to reward because deception has greater consequences.

Our experiments equated costs and outcomes in three ways: Costs of punishing and rewarding were symmetric, participants could not compare their payoffs with those of their counterparts, and deception hurt as much as honesty helped. This last methodological improvement required two honesty conditions. One was structured such that honest and deceptive actions affected participants' outcomes by equal absolute amounts; for example, people gained $50 from honesty (i.e., they expected $50 and received $100) or lost $50 from deception (i.e., they expected $150 and received $100). The other honesty condition was structured such that deception and honesty had equal relative effects on participants' wealth (Thaler, 1980). For example, if participants expected $150 but received $100 because of deception, they might view deception as causing a 33% loss. Therefore, we designed the corresponding honesty condition with relative effects equated so that participants would expect $75 (rather than $50), so that when they received $100, they experienced a 33% (rather than 100%) gain. Thus, we compared reactions to deception with reactions to honesty both when the gains and losses were equal in absolute terms and when they were equal in relative terms.

Preview of the Experiments

Experiment 1 tested people's reactions to honesty and deception, and we found that people rewarded honesty more than they punished deception. In Experiment 2, we investigated a potential divergence between judgment and behavior (the possibility that deception would produce more extreme evaluations than honesty, but honesty would produce stronger behavioral intentions than deception) and compared the degree to which people approached honest actors versus avoided deceivers. This study provides an empirical basis for our theoretical reconciliation with the "bad is stronger than good" literature. In Experiment 3, we again observed more rewarding than punishing, even when participants incurred real costs in punishing or rewarding their counterparts. We also found different psychological mechanisms for punishing and rewarding, with negative affect driving punishment and cognitive trust driving rewards.

EXPERIMENT 1

Method

Participants were 104 undergraduates (44 men, 60 women). Experiment 1 had a 3 (behavior: deception, honesty with absolute outcomes equated, honesty with relative effects equated) × 2 (cost: low, high) between-subjects factorial design.

Participants read about someone who behaved dishonestly or honestly. Instructions in the dishonesty condition read: "Imagine the following scenario: You and Pat recently completed a

business deal; you have just discovered that Pat was dishonest about some key information. As a result, you only received $100. You would have received 50% more if Pat had been honest." Instructions in the honesty condition with absolute outcomes equated read: "Imagine the following scenario: You and Pat recently completed a business deal; you have just discovered that Pat was honest about some key information. As a result, you received $100. You would have received 50% less if Pat had been dishonest."

Thus, participants in both of these conditions were told they received $100. In the dishonesty condition, they were told they experienced an absolute loss of $50, equivalent to a 33% relative loss. In the honesty condition with absolute outcomes equated, they were told they experienced an absolute gain of $50, equivalent to a 100% relative gain. Participants in the honesty condition with relative effects equated read a similar scenario and were also told they received $100. However, they were told they would have received $75 without Pat's honesty; thus, they experienced an absolute gain of $25, equivalent to a 33% relative gain.

Next, participants were given the opportunity to spend hypothetical money to reward (in the honesty conditions) or punish (in the dishonesty condition) Pat. In the high-cost condition, participants paid out exactly the amount of reward or punishment they chose. In the low-cost condition, costs were only 10% of the reward or punishment amount (Brandts & Charness, 2003). For example, in the low-cost, dishonesty condition, the instructions read:

> You have a one-time opportunity to punish Pat, but responding will cost you money. You can choose to behave in a fashion that is equivalent to taking money from Pat at a 1:10 ratio: 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.

Punishment and reward amounts, along with their costs, were presented on 11-point scales from $0 to $100, in $10 increments. For example, in the high-cost, dishonesty condition, the option for a punishment of $10 read, "Punish $10 (at a cost of $10)." The two dependent measures were the *frequency* of rewards and punishments and their *intensity* (dollar amount).

Results and Discussion

Women rewarded or punished Pat more often than men (77% vs. 59%), $\chi^2(1, N = 104) = 4.60$, $p < .05$, and also chose larger rewards and punishments than men did ($M = 46.92$, $SD = 38.28$, vs. $M = 33.41$, $SD = 36.21$), $F(1, 102) = 2.81$, $p < .10$ (Croson & Buchan, 1999). Therefore, we report results of analyses controlling for gender.

Frequency

Figure 1 shows the percentage of participants who decided to reward or punish Pat in each condition. Binomial logistic regression analyses, with deception as the reference category (42%), indicated that rewards were approximately twice as frequent as punishments, both when conditions were equated in absolute terms (89%), $b = 3.00$, $SE = 0.93$, $Wald = 10.36$, $p < .01$, and when conditions were equated in relative terms (74%), $b = 1.98$, $SE = 0.86$, $Wald = 5.27$, $p < .05$. Frequencies in the two honesty conditions did not differ significantly, $b = -1.27$, $SE = 2.58$, $Wald = 0.24$, $p = .62$. Cost had a marginal effect,

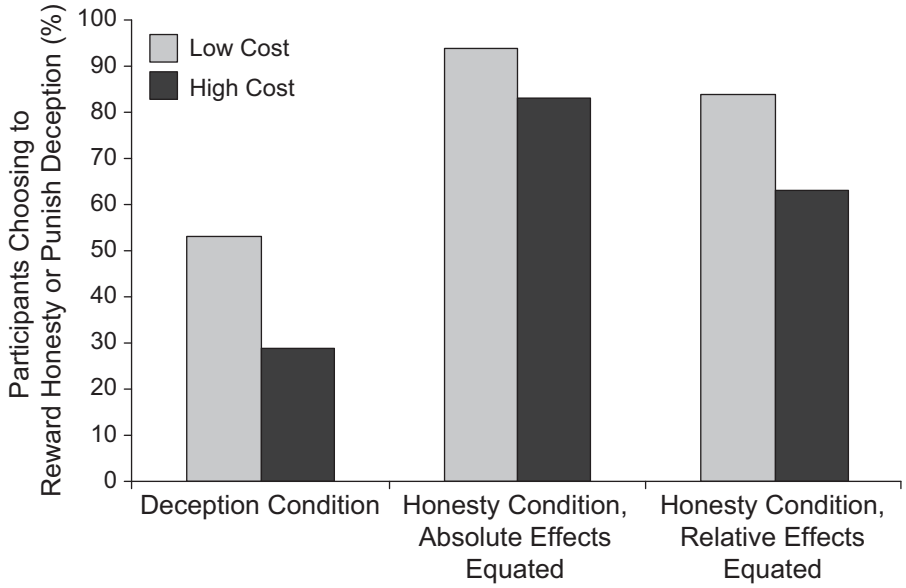

Fig. 1. Results from Experiment 1: percentage of participants rewarding or punishing a business partner as a function of behavior condition (deception, honesty with absolute outcomes equated, and honesty with relative outcomes equated) and cost to administer the reward or punishment (high or low).

$b = 1.49$, $SE = 0.81$, $Wald = 3.39$, $p < .07$; participants were more likely to reward or punish Pat when costs were low than when they were high (77% vs. 60%, respectively). The interaction between behavior (coded as a categorical variable) and cost (coded as a continuous variable) was not significant, $Wald = 0.29$, $p = .87$.

Intensity

Rewards were also larger than punishments, both when conditions were equated in absolute terms ($M = 48.89$, $SD = 28.86$, vs. $M = 28.18$, $SD = 39.33$), $F(1, 97) = 8.71$, $p < .01$, and when they were equated in relative terms ($M = 45.57$, $SD = 42.16$, vs. $M = 28.18$, $SD = 39.33$), $F(1, 97) = 5.74$, $p = .02$. The amount of reward did not differ between the two honesty conditions, $F < 1$. Low costs led to significantly more intense responses ($M = 57.41$, $SD = 40.31$) than high costs ($M = 22.29$, $SD = 23.63$), $F(1, 97) = 35.01$, $p < .001$. The interaction of behavior and cost was not significant, $F(2, 97) = 0.14$, $p = .87$ (see Fig. 2).

Discussion

Because Experiment 1 equalized both the costs to punish and to reward an actor and the consequences of deception and honesty, absolutely and relatively, it provides a clear comparison of individuals' reactions to honesty and deception. Unlike in previous studies, participants rewarded honesty more frequently and intensely than they punished deception. Although they also rewarded honesty and punished deception less frequently and intensely when costs were high, rather than low, costs did not moderate the tendency for rewards to be more frequent and intense than punishments.

EXPERIMENT 2

Our theoretical reconciliation regarding the relative strengths of good and bad suggests the possibility of an asymmetry between evaluation and behavior: People might dislike deceivers more than they like truth tellers, but punish deceivers less than they reward truth tellers. Other evidence suggests that evaluation and behavior can be independent (Galinsky, Ku, & Wang, 2008; Mussweiler & Förster, 2000). For example, compared with nonperspective takers, perspective takers judge other people less stereotypically (Galinsky & Moskowitz, 2000), but behave more stereotypically themselves (Galinsky, Wang, & Ku, 2008). In Experiment 2, we investigated whether deception leads to stronger psychological reactions but weaker behavioral responses than honesty.

Participants in Experiment 1 may have chosen to reward more than to punish because desirable events encourage approach and negative events encourage avoidance (James, 1890). Indeed, people are reluctant to transmit bad news and are more willing to transmit good news (i.e., the tendency to keep mum about undesirable messages, or the MUM effect; Tesser & Rosen, 1975). Thus, rewards may be favored over punishments because deception engenders competing motivations to avoid versus to punish, whereas honesty engenders complementary motivations to approach and to reward (Eisenberger et al., 2004;

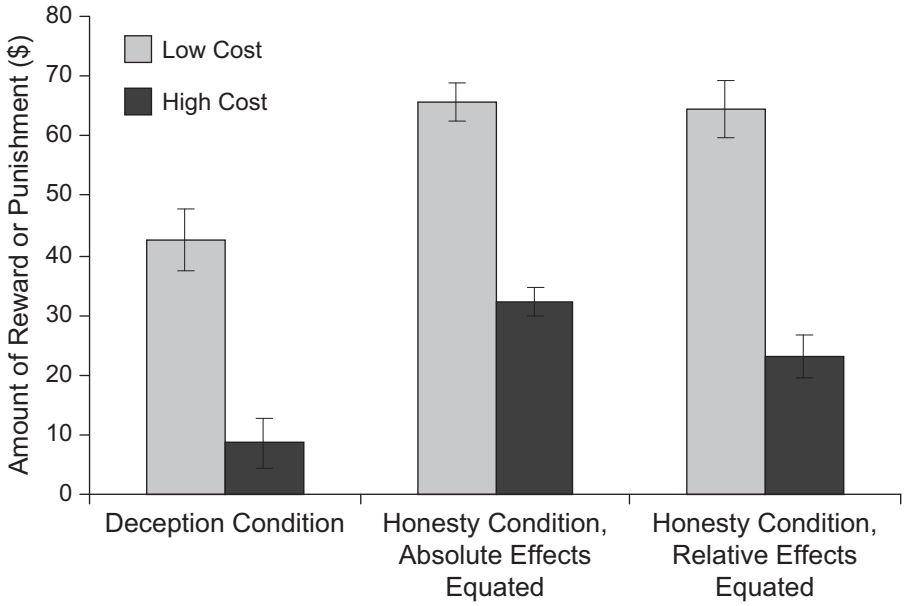

Fig. 2. Results from Experiment 1: mean amount of reward or punishment as a function of behavior condition (deception, honesty with absolute outcomes equated, and honesty with relative outcomes equated) and cost to administer the reward or punishment (high or low). Error bars indicate standard errors of the means.

Hokanson, 1974). We also investigated this possibility in Experiment 2.

Method

Participants were 64 undergraduates (25 men, 39 women). Experiment 2 had a 2 (behavior: deception, honesty with absolute outcomes equated) × 2 (measure: evaluation, behavior) factorial design. The second factor was manipulated within participants.

The procedure was identical to that of Experiment 1 except that costs were constant at a ratio of 1:10 (i.e., low cost) and we included only one of the honesty conditions—the honesty condition with absolute effects equated—because results for the two honesty conditions did not differ in Experiment 1. In addition, participants not only decided whether to reward or punish Pat, but also evaluated Pat on four 100-point scales (in the deception condition: unfair, untrustworthy, risky, and negative; in the honesty condition: fair, trustworthy, safe, and positive). For example, in the deception condition, participants read, "Indicate how unfair you perceive the individual to be using a scale between 0 (least unfair) and 100 (most unfair)." Given high reliability, we averaged across the four measures in each condition ($\alpha = .91$ for the honesty dimensions and .74 for the deception dimensions; Falk & Fischbacher, 2006); higher numbers represented more extreme evaluations: more negative in the deception condition and more positive in the honesty condition. Participants' feelings of approach-avoidance were assessed by the following item: "I would[honesty condition]/would not[deception condition] like to spend time with this individual"; the response scale ranged from 1 (*extremely disagree*) to 10 (*extremely agree*). Thus, higher scores indicate greater approach in the honesty condition and greater avoidance in the dishonesty condition.

The order of the punishment/reward and evaluation measures was counterbalanced. Order and gender did not affect the results, so we collapsed across these factors.

Results and Discussion

We expected negative evaluations to be stronger than positive evaluations, but rewards to be more frequent and intense than punishments. Given that both the behavior and the evaluation items used scales from 0 to 100, we conducted a 2 (behavior) × 2 (measure) mixed-model analysis of variance, which revealed a significant interaction, $F(1, 62) = 13.93$, $p < .001$.

Participants' behavioral responses replicated those in Experiment 1: Rewards (97%) were more frequent than punishments (65%), $b = 2.87$, $SE = 1.08$, $Wald = 7.02$, $p < .01$, and were also more intense than punishments, $t(63) = 8.27$, $p < .01$ (see Fig. 3).

Participants' evaluations, however, revealed the reverse pattern: Participants judged deceptive actors more negatively than they judged honest actors positively, $t(63) = 3.84$, $p = .05$ (see Fig. 3). Thus, participants exhibited stronger evaluative reactions toward deception than toward honesty, but stronger behavioral intentions toward honesty than toward deception.

In addition, participants expressed a greater desire to avoid a deceiver ($M = 7.39$, $SD = 2.82$) than to approach an honest person ($M = 5.24$, $SD = 2.14$), $t(63) = 11.83$, $p = .001$. These findings demonstrate a divergence between evaluation and behavior while also suggesting that people choose to reward more than to punish because of strong motivations to avoid deceivers.

EXPERIMENT 3

The next experiment went beyond the hypothetical nature of the first two experiments by including behavioral responses with real monetary consequences. It also investigated two underlying psychological mechanisms: perceptions of trustworthiness and negative affect.

People typically reciprocate trustworthiness (Pillutla, Malhotra, & Murnighan, 2003; Tesser, Gatewood, & Driver, 1968). In addition, neurological data suggest that people have positive neural reactions to partners who have been consistently trustworthy (King-Casas et al., 2005). We predicted that perceived trustworthiness would be the driving force behind rewarding honesty.

Evidence also suggests that negative affect drives even costly punishments (Loewenstein, 1996). Pillutla and Murnighan (1996) found that anger at receiving small offers in an ultimatum game was the primary cause for rejecting these offers, and brain activity in the area associated with negative affect increases when responders reject low, unequal ultimatum offers (e.g., less than a \$5-\$5 split; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). We predicted that when people punished dishonesty, negative affect would be the generative cause.

Method

Participants were 176 undergraduates (79 men, 97 women). Experiment 3 had a 3 (behavior: deception, honesty with absolute outcomes equated, honesty with relative effects equated) × 3 (cost: none, low, high) between-subjects factorial design. Participants received \$5 for participating, in addition to any money they accumulated in the experiment.

Participants were informed that they would interact with another player who could act honestly or deceptively in a two-stage game in which they would be deciding between two options, A and B (modified from Gneezy, 2005). In Stage 1, the other player could ostensibly send participants a true or an untrue message that would read, "Option A earns you more than Option B" or "Option B earns you more than Option A." In fact, all participants received the former message. They subsequently chose between Options A and B and then received feedback about whether the message had been honest or deceptive. The final payoff was held constant (\$4) in all conditions.

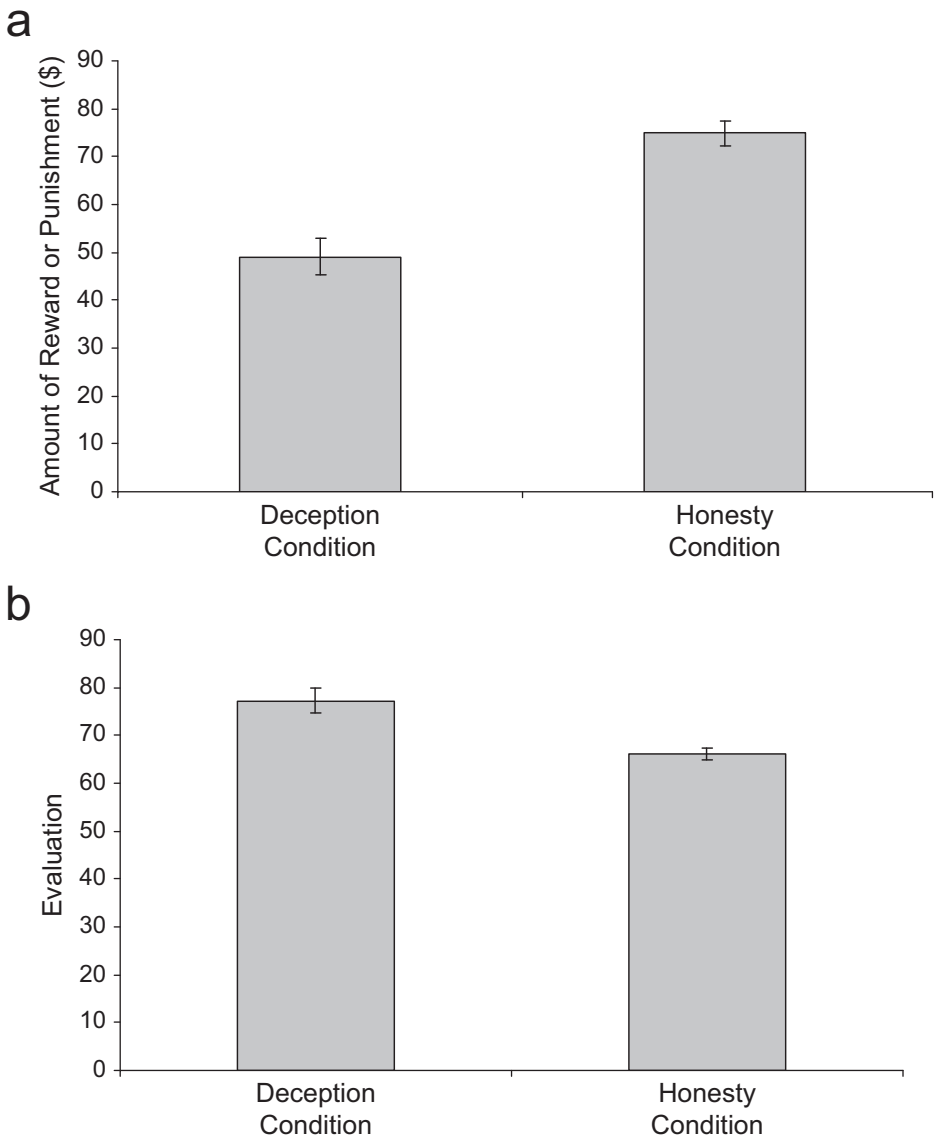

Fig. 3. Results from Experiment 2: (a) mean amount of reward or punishment and (b) mean evaluation of the partner as a function of whether the partner had been deceptive or honest. Error bars indicate standard errors of the means. In (b), the bars indicate the level of positive evaluations in the honesty condition and negative evaluations in the deception condition.

In the deception condition, participants who chose Option A (which suggests that they believed the message) were told: "We can now reveal that the true message was: Option B will earn you more than Option A. Because you chose Option A, you receive $4 in Stage 1. You would have received 50% more if you had chosen Option B." Participants who chose Option B saw a different ending: "Because you chose Option B, you receive $4 in Stage 1. You would have received 50% less if you had chosen Option A."

In the honesty condition with absolute outcomes equated, participants who chose Option A were told: "We can reveal that the true message was: Option A will earn you more than Option B. Because you chose Option A, you receive $4 in Stage 1. You would have received 50% less if you had chosen Option B." Participants who chose Option B saw a different ending: "Because you chose Option B, you receive $4 in Stage 1. You would have received 50% more if you had chosen Option A." The honesty condition with relative effects equated was identical except that participants were told they would have received 25% less (instead of 50% less) when they believed the message and 25% more when they did not believe it.

In Stage 2, participants could use their $4 payoff to reward or punish the other player. In the low-cost condition, the cost was 10% of the amount of the punishment, and in the high-cost condition, it was equal to the amount of the punishment or reward. In the no-cost condition, participants could punish or

reward the other player without using their $4 payoff. We then measured perceptions of the other player's trustworthiness and participants' negative affect. Perceptions of trustworthiness were assessed with four items adapted from Pillutla et al. (2003; $\alpha = .88$). For example, the item "I consider the other player to be . . ." had response options ranging from *completely untrustworthy* (1) to *completely trustworthy* (7). Six items assessed negative affect ($\alpha = .82$). For example, one item asked participants how they felt after finding out about the other player's deception or honesty; response options ranged from *completely happy* (1) to *completely unhappy* (7).

All participants received $9 minus the cost of any reward or punishment they had given.

Results

Women chose to reward or punish the other player marginally more frequently than men did (73% vs. 61%), $\chi^2(1, N = 176) = 3.08$, $p < .08$, and also chose marginally larger rewards and punishments than men did ($M = 1.92$, $SD = 1.56$, vs. $M = 1.48$, $SD = 1.49$), $F(1, 97) = 2.81$, $p < .10$ (Croson & Buchan, 1999). Therefore, we report results of analyses controlling for gender. Whether or not participants believed the other player did not have a significant main effect and was not involved in any significant interactions (all Fs < 1.4); as a result, we did not control for variations in whether or not they followed the other party's advice.

Frequency

Participants chose to reward more often than to punish (35%), both when conditions were equated in absolute terms (87%), $b = 1.71$, $SE = 0.56$, $Wald = 9.41$, $p < .01$, and when conditions were equated in relative terms (92%), $b = 3.12$, $SE = 1.17$, $Wald = 7.09$, $p < .01$. As before, the frequency of choosing to reward did not differ between the two honesty conditions, $b = -1.41$, $SE = 1.20$, $Wald = 1.37$, $p = .24$ (see Fig. 4).

A significant interaction between behavior and cost (with behavior as a categorical variable and cost as a continuous variable), $Wald = 6.01$, $p = .05$, suggested that punishments were less frequent as costs increased (see Fig. 4). Specifically, frequency of punishments differed significantly between the high-cost and no-cost deception conditions, $\chi^2(1, N = 49) = 5.68$, $p < .05$, and also between the high-cost and low-cost deception conditions, $\chi^2(1, N = 49) = 13.91$, $p < .001$; frequency of punishments did not differ significantly between the no-cost and low-cost deception conditions, $\chi^2(1, N = 40) < 1$. The frequency of rewards in the honesty conditions, in contrast, was essentially unaffected by costs, all χ^2s < 1.

Intensity

People also rewarded more intensely than they punished, both when conditions were equated in absolute terms, $F(1, 167) = 25.22$, $p < .001$, and when they were equated in relative terms, $F(1, 167) = 18.83$, $p < .001$; intensity of reward did not differ significantly between the two honesty conditions, $F < 1$. Levels of punishment and reward in the high-cost condition ($M = 1.00$, $SD = 1.15$) were significantly lower than those in the no-cost condition, $F(1, 167) = 10.32$, $p < .01$, and the low-cost condition, $F(1, 167) = 12.72$, $p < .001$, which did not differ significantly ($M = 2.10$, $SD = 1.61$, and $M = 2.11$, $SD = 1.57$, respectively). The interaction of behavior and cost was not significant, $F(4, 167) < 1$ (see Fig. 5).

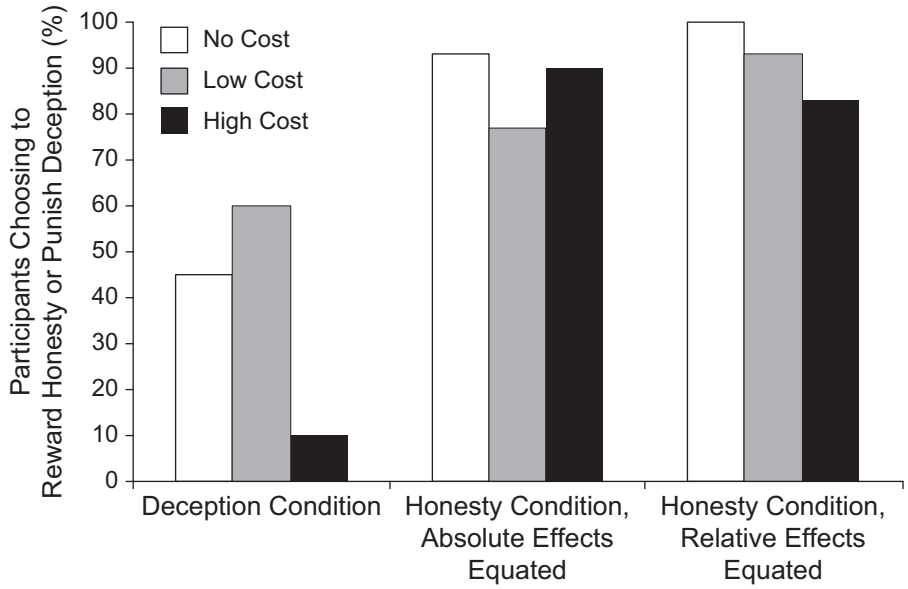

Fig. 4. Results from Experiment 3: percentage of participants rewarding or punishing the other player as a function of behavior condition (deception, honesty with absolute outcomes equated, and honesty with relative outcomes equated) and cost to administer the reward or punishment (high cost, low cost, or no cost).

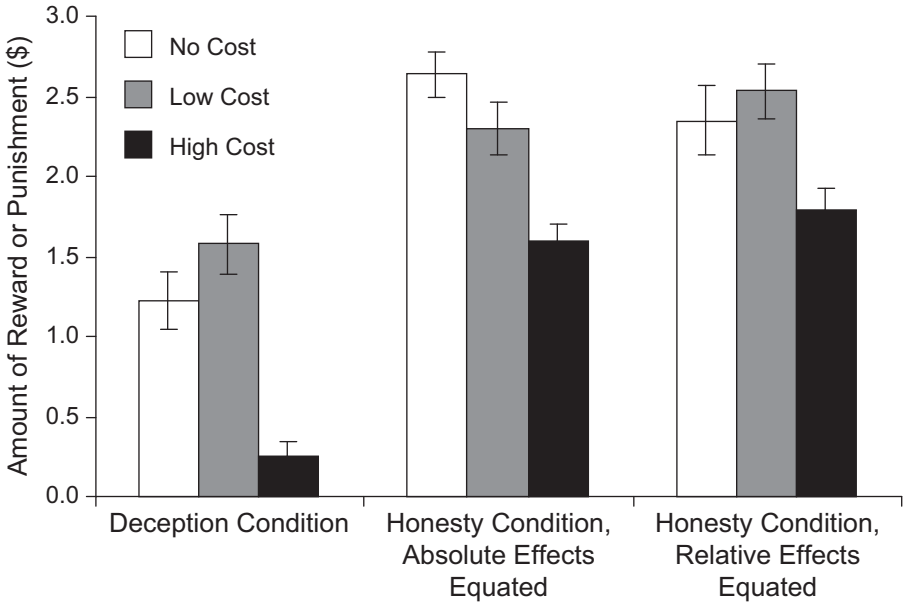

Fig. 5. Results from Experiment 3: mean amount of reward or punishment as a function of behavior condition (deception, honesty with absolute outcomes equated, and honesty with relative outcomes equated) and cost to administer the reward or punishment (high cost, low cost, or no cost). Error bars indicate standard errors of the means.

Perceived Trustworthiness and Negative Affect

Participants rated honest players as more trustworthy than dishonest players ($M = 3.38$, $SD = 1.01$), both when conditions were equated in absolute terms ($M = 5.25$, $SD = 0.89$), $F(1, 167) = 137.53$, $p < .001$, and when they were equated in relative terms ($M = 5.37$, $SD = 0.87$), $F(1, 167) = 113.18$, $p < .001$.

Participants reported more negative affect following deception ($M = 4.01$, $SD = 0.84$) than following honesty (absolute condition: $M = 2.87$, $SD = 0.99$; relative condition: $M = 2.98$, $SD = 0.93$). This difference was significant both when conditions were equated in absolute terms, $F(1, 167) = 47.73$, $p < .001$, and when they were equated in relative terms, $F(1, 167) = 30.80$, $p < .001$.

Moderated-Mediation Analyses of Intensity

We conducted two moderated-mediation analyses (see Preacher, Rucker, & Hayes, 2007) to test the impact of negative affect and perceived trustworthiness on punishment and reward amounts. Moderated mediation occurs when the mediator and independent variable interact to cause the outcome (Muller, Judd, & Yzerbyt, 2005).

We predicted that perceived trustworthiness would mediate the relationship between honesty and reward, but not the relationship between deception and punishment, such that perceptions of trustworthiness would positively predict reward amounts, but would have no relationship with punishment amounts. We also predicted that negative affect would mediate the relationship between deception and punishment, such that negative affect would positively predict punishment amounts, but would have no relationship with reward amounts.

To test the impact of perceived trustworthiness, we conducted a regression that included behavior (honesty/deception manipulation), trustworthiness perceptions, and the Trustworthy × Behavior interaction (reflecting a moderation effect), with cost, gender, and negative affect as covariates. The two honesty conditions were combined because intensity did not differ between these conditions. The trustworthy-behavior interaction was significant, $\beta = .56$, $SE = .22$, $p < .02$. This result indicates that the association between perceived trustworthiness and intensity depended on the other player's behavior. Simple slopes of the interaction (Holmbeck, 2002) showed that, as predicted, perceptions of trustworthiness predicted reward amounts, but were not related to punishment amounts (see Fig. 6a).

A comparable analysis involving negative affect revealed a significant Negative Affect × Behavior interaction, $\beta = -.67$, $SE = .23$, $p < .01$. Simple slopes demonstrated that greater negative affect significantly increased punishments, but marginally decreased rewards (see Fig. 6b). This finding partially supported our prediction for negative affect.

In summary, whereas perceived trustworthiness mediated the honesty-reward relationship, negative affect mediated the deception-punishment relationship.

GENERAL DISCUSSION

Across three experiments, individuals rewarded honesty more than they punished deception. Given that we both held constant the impact of honesty and deception and equalized the costs of

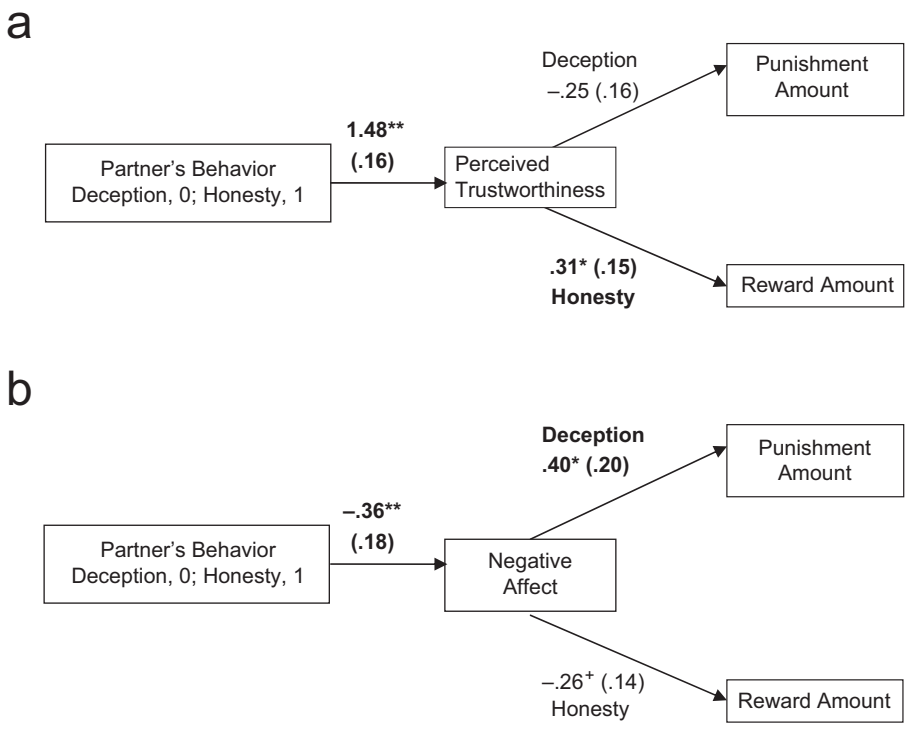

Fig. 6. Results of the moderated-mediation analyses in Experiment 3: analyses of the effect of (a) perceived trustworthiness and (b) negative affect on the amount of reward and punishment. The figure shows *b* coefficients, with standard errors in parentheses, $^{+}p \leq .10$, $^{*}p \leq .05$, $^{**}p < .001$. Boldface highlights coefficients indicating significant mediation.

rewards and punishments, while using multiple methods, contexts, and operationalizations, these experiments pose a strong challenge to previous findings that individuals punish deception more than they reward honesty (Abbink et al., 2000; Brandts & Charness, 2003; Offerman, 2002).

Further Theoretical Reconciliation

In combination with past findings (i.e., the MUM effect, Tesser & Rosen, 1975), the current findings suggest that bad affects evaluations more than good does, but that good affects behavior more than bad does. These patterns may ultimately be explained by the relative psychological impact of reward and punishment: People may be reticent to administer punishments because they are aware that punishment has stronger psychological effects than reward, and they may feel that they need to distribute rewards more frequently than punishments in order for rewards to have an impact (Baumeister et al., 2001).

Our results also suggest that socially excluding transgressors may be preferred over direct punishment. Because social contact is fundamental to well-being (Baumeister & Leary, 1995), avoidance can serve as an indirect but potent form of punishment. Of course, avoidance may be geared less toward punishing and more toward future-oriented self-protection. Both retrospective punishment and prospective protection may lead to avoidance rather than punishment of deceivers. Although Experiment 3 indicates that punishments are more sensitive to costs than rewards are, negative emotions like anger, when strong enough, may push people past their avoidance tendencies, releasing them to punish even when that requires significant sacrifice.

This reticence to directly punish deceivers may tacitly reward small, negative actions that develop into increasing wrongdoing. As Sherron Watkins, the Enron whistle-blower, noted, "Star performers that violate the company's procedures are too often given a second chance" (Gleason, 2006).

Experiment 3 suggests that the dynamics of punishment and reward are not mirror images of one another. Rather, reward and punishment are activated by different cognitive and affective mechanisms that ultimately lead to different behavioral choices; bad begets bad, and good stimulates good, but through different pathways. Whereas negative affect drives punishment, rewards travel on a different highway, via perceived trustworthiness.

Limitations and Future Considerations

Future research should clarify the role of certain methodological components held constant in this study. In our experiments, participants may have assumed that individuals sacrificed their own money to be honest; future studies should test whether people will reward honesty from truth tellers who have monetarily benefited from their honesty. Another issue concerns ex-

pectancy violations: Future research could test people's reactions when they expect an amount that they then actually receive. Our results suggest that expectations and their violations may affect rewards and punishments less than one might think: Reward behavior did not differ between the two honesty conditions (e.g., 100% vs. 33% absolute gain), and participants' choices to reward and punish were similar regardless of expectation violation or confirmation (e.g., in Experiment 3, punishments did not differ between participants who believed the other player's message and received 50% less than expected and those who did not believe the message and received 50% more than expected).

Finally, future research should explore other mediators. One potentially relevant psychological state, inherently tied to perceived trustworthiness, is felt obligation. Indeed, Wang (2007) found that participants rewarded honesty more than they punished deception because honesty induced an obligation to reciprocate.

CONCLUSION

The current results provide a synthesis of the economic and social psychological literatures on rewarding honesty and punishing deception. The result is a new perspective on the "bad is greater than good" argument, one that warrants additional empirical research and richer theoretical conceptualizations on how and why people choose to punish and to reward.

Acknowledgments—This research was based in part on the doctoral dissertation submitted by Cynthia Wang to Northwestern University and has benefited from the comments of committee members David Austen-Smith and Katherine Phillips. We thank Katie Dover-Taylor, Laura Howland, and Will Thomson for their research assistance and Chris Bauman, Deborah-Anna Reznek, Niro Sivanathan, Leigh Thompson, and Jennifer Whitson for their helpful comments. This work benefited from the generous financial support by the Dispute Resolution Research Center at the Kellogg School of Management.

REFERENCES

Abbink, K., Irlenbusch, B., & Renner, E. (2000). The moonlighting game: An experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization*, *42*, 265–277.

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., & Vohs, K.D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 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.

Brandts, J., & Charness, G. (2003). Truth or consequences: An experiment. *Management Science*, *49*, 116–130.

Cropanzano, R., & Mitchell, M.S. (2005). Social exchange theory: An interdisciplinary review. *Journal of Management*, *31*, 874–900.

Croson, R., & Buchan, N. (1999). Gender and culture: International experimental evidence from trust games. *American Economic Review*, *89*, 386–391.

Dreben, E.K., Fiske, S.T., & Hastie, R. (1979). The independence of evaluative and item information: Impression and recall order effects in behavior-based impression formation. *Journal of Personality and Social Psychology*, *37*, 1758–1768.

Eisenberger, R., Lynch, P., Aselage, J., & Rohdieck, S. (2004). Who takes the most revenge? Individual differences in negative reciprocity norm endorsement. *Personality and Social Psychology Bulletin*, *30*, 789–799.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*, 293–315.

Fehr, E., & Gachter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*, 159–181.

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.

Galinsky, A.D., Ku, G., & Wang, C.S. (2008). *Perception through a perspective-taking lens: Differential effects on judgment and behavior*. Unpublished manuscript, Northwestern University, Evanston, IL.

Galinsky, A.D., & Moskowitz, G.B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and ingroup favoritism. *Journal of Personality and Social Psychology*, *78*, 708–724.

Galinsky, A.D., Wang, C.S., & Ku, G. (2008). Perspective-takers behave more stereotypically. *Journal of Personality and Social Psychology*, *95*, 404–419.

Gleason, R. (2006, November 16). Enron whistleblower tells tale of corruption. *The Houstonian*. Retrieved March 3, 2009, from http://www.houstonianonline.com

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*, 384–394.

Gouldner, A.W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, *25*, 161–178.

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*, 87–96.

James, W. (1890). *The principles of psychology* (Vol. 2). New York: Holt.

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*, 78–83.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*, 272–292.

Muller, D., Judd, C.M., & Yzerbyt, V.Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852–863.

Mussweiler, T., & Förster, J. (2000). The sex-aggression link: A perception-behavior dissociation. *Journal of Personality and Social Psychology*, *79*, 507–520.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, *46*, 1423–1437.

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.

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

Pillutla, M.M., & Murnighan, J.K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, *68*, 17.

Preacher, K.J., Rucker, D.D., & Hayes, A.F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*.

Rozin, P., & Royzman, E.B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 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*, 1755–1758.

Tesser, A., Gatewood, R., & Driver, M. (1968). Some determinants of gratitude. *Journal of Personality and Social Psychology*, *9*, 233–236.

Tesser, A., & Rosen, S. (1975). The reluctance to transmit bad news. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8, pp. 193–232). New York: Academic Press.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*, 39–60.

Wang, C.S. (2007). *Rewarding honesty and punishing deception*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

(Received 8/26/08; Revision accepted 10/25/08)