The Road to Hell: An Experimental Study of Intentions

Gary Charness and David I. Levine*

July 19, 2004

Abstract: Do people care about intentions – even when good intentions do not produce good results? While the aphorism in the paper's title suggests that outcomes matter more than intentions, our study questions that assumption.

Our experimental design builds on previous studies of intentions by introducing an element of chance in the wage-determination process in an experimental labor market. Rates of punishment and reward react strongly to intentions (the wage a firm decides to pay) and modestly to distributional outcomes (the higher or lower wage actually received after chance intervenes). For example, workers who receive "medium" wages are much kinder to a boss who tried to pay a high wage (but bad luck lowered the worker's pay) than they are to a boss who tried to pay a low wage (and good luck raised pay).

Keywords: Intentions, Reciprocity, Experiment, Rent-sharing, Process, Attribution

JEL Classification: A13, B49, C91, D63, J24, J41

* Charness: Dept. of Economics, UCSB. Email charness@econ.ucsb.edu. Levine: Haas School of Business, U.C. Berkeley. Email: levine@haas.berkeley.edu. We appreciate help from Brian Wei in conducting experimental sessions, and acknowledge helpful suggestions from Matthew Rabin and Barbara Mellers.

Even a dog knows the difference between being stepped on and being tripped over.
-- Oliver Wendell Holmes

Outcomes for any given situation often depend on a tricky combination of intentional effort and luck. To assess the importance of good intentions in the workplace, we ask: Under what circumstances do people pay attention to outcomes, and under what circumstances do they focus on intentions? Considerable evidence indicates that monetary reward is not the only motivation present among economic agents; social preferences such as altruism and reciprocity play roles as well.

From Adam's (1965) classic equity theory to recent economic models of fairness in games, many social scientists have extended the assumption of self-interested preferences to include the idea that judgments and actions reflecting fairness are based on relative outcomes (see Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). These distinctions occur in law, as suggested by the opening quote from Oliver Wendell Holmes and as argued by Huang (2000). For example, manslaughter receives a lighter sentence than second-degree murder, which in turn receives a lighter sentence than first-degree murder. All of these crimes result in a deceased victim, but they differ in the perpetrator's perceived intentions.

Kahneman, Knetsch, and Thaler (1986) and Charness and Levine (2002) find that people's perceptions of fairness in consumer markets and the workplace differ for identical actions depending upon whether external circumstances provide a good justification for the action. Bewley notes that because morale is important for workplace performance, "reciprocity and even humanitarian feelings have an impact on behavior" (1999: 56). Employees' differences in perceptions about whether particular wages are justified may well lead to differences in

performance. A worker who feels unfairly treated by his or her employer will be less likely to feel loyal to the firm when choosing effort.

The assumption that consequences matter but that processes do not has been challenged both theoretically and experimentally. Economic models of kindness-based reciprocity offer perceived kindness or unkindness as the primary motivation for why people choose not to maximize their own material payoffs (see Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). The more recent models proposed by Falk and Fischbacher (forthcoming) and Charness and Rabin (2002) enrich the outcome-based models by incorporating both perceived intentions and distributional concerns and reactions. Several experiments find support for the notion that intentions matter with regard to economic choices. However, several other experiments do not find that perceived intentions affect behavior.

Previous experimental studies of intention rely upon comparisons that may be confounded. One such approach compares responses made to an intentional choice by a first mover to responses to the same choice when it is produced by exogenous conditions such as a draw from a bingo cage. This approach has limitations, however; for example, some research suggests that people may be more generous when the affected party has no choice in the process, in which case such a comparison may not isolate the effect of perceived intention.[1] In another approach, participants indicate contingent responses to each of a large number of feasible interim outcomes. While this methodology may seem innocuous, evidence indicates that a within-subject design often leads to different results than does the more standard between-subjects approach.

[1] For evidence that a person may be more generous when the other party has no choice about or responsibility for the interim outcome, see Charness (2000), Charness and Rabin (2002), and Morgenstern (2004).

We provide a clean test of intention's influence in an experimental employment relationship by explicitly separating intentions from outcomes. Our study includes an element of chance in the wage-determination process and also permits workers to sacrifice to either help or hurt the firm. We permit a particular outcome to arise from either: (1) good intentions (a high wage costly to the firm) coupled with bad luck, or (2) less-good intentions (a less-costly low wage) coupled with good luck. If only outcomes matter, the employees' responses should be similar regardless of how this situation arose. If intentions matter, employees should be more likely to work hard and less likely to sabotage the employer when the employer's intentions were good.

We find that the rates of punishment and reward are sensitive to both the wage selected by the employer and the amount actually received by the employee after Nature has intervened. Despite identical wages and identical relative payoffs, responders' behaviors differ strongly. The motivations underlying non-self-interested behavior have important economic applications, including consumer response to price changes, attitudes toward different tax schemes, and employee response to changes in wage and employment practices.[2]

Background & Previous Work

While the standard neoclassical economic model assumes that people care only about maximizing their own payoffs, other social sciences have long considered the role played by social forces. Procedural justice theories emphasize people's concern about whether decision-makers seem fair and respectful or self-interested and disrespectful (e.g., Tyler, 1988); field evidence suggests that when the latter attributions predominate, people are more inclined to quit,

work less hard, and be generally less productive. Levine (1993) concludes that an employee who feels he or she is receiving fair treatment is more likely to perform above any minimum requirement.

Adams (1965) is concerned with equity---the fair share deserved by the constituent members of a group--as determined by the relative inputs of the group members.[3] Rent-sharing models from labor economics also express this view (at least implicitly), with workers and managers bargaining over a pie of quasi-rents. In such models, exogenously good or bad business conditions lead to high or low wages (Blanchflower, Oswald, and Garrett, 1990; Bertrand and Mullainathan, 1999).

People may also care about the process that led to payoffs, not just their own and others' material payoffs. Heider (1958) points out the importance of casual inference, where a person takes into account another actor's motive and situational constraints, as a cognitive process for perceiving social context. Gouldner (1960) argues that responses depend on the initial agent's perceived motives and free will.

A key issue is how people determine attributions for observed outcomes; often people will conflate outcomes and abilities. The fundamental attribution error reveals this (Ross, 1977): In many situations, people attribute more responsibility for outcomes to a person (rather than to a situation) than is warranted. Attribution theory considers a chain of events in which a person asks why an outcome occurred, assigns an attribution for the cause, and behaves accordingly (see Nisbett and Ross, 1980). Many psychology experiments support this notion: Greenberg and Frisch (1972) find that help that is deliberately given leads to more reciprocity than does

[2] Fehr and Gächter (2000) provide a summary of economic applications for fairness and reciprocity.
[3] This concept is reflected in the Rawlsian utility function, in which the society is only as well off as its least well-off member.

accidental help.[4] In Blount (1995), responders in an ultimatum game will accept a substantially lower proposal when it is generated randomly than when chosen by a self-interested party.

Economists have not yet generally grappled with the value of intentions. For example, Waldfogel (1993) estimates that Christmas gift-giving leads to a deadweight loss in the billions of dollars. To estimate this figure, he compares self-reports of gifts' values--*not counting sentimental value*--with the recipient's estimates of their market prices. He then estimates that gift-giving "destroys" between a tenth and a third of the gifts' value (p. 1336).[5] Yet, if a gift's sentimental value matters ("It's the thought that counts…"), then holiday gifts may create value, even if the recipients would not have chosen these specific gifts.

In the past decade, these views have inspired behavioral-economic models of utility. The Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) models assume that people like money but dislike inequality in material payoffs. Thus, agents may sacrifice their own payoff to reduce disparities in relative payoffs. The Fehr and Schmidt model also presumes that people who draw the short stick may dislike inequality more than those who draw the long one.

Another related set of models assumes that people like to maximize total payoffs yet are particularly concerned about the level of the lowest payoff that any player receives. These models have substantial experimental support (e.g., Frohlich and Oppenheimer 1984, 1992; Charness, 2004; Andreoni and Miller, 2002; Kritikos and Bolle, 1999; Charness and Grosskopf, 2001; Charness and Rabin, 2002). In the two-player games we study, concern about the lowest payoff is indistinguishable from concern about inequality.

[4] See also Thibaut and Riecken (1955), Goranson and Berkowitz (1966), Kelley and Stahelski (1970), Kelley (1972), and Kahn and Tice (1973).

[5] In later comments, he notes that people may value gift-giving (because of the thought involved), even if they do not give much value to the gift objects (Waldfogel 1996: 1306). Nevertheless, he stands by his description of gift-giving as a producer of deadweight loss.

Rabin (1993) and Dufwenberg and Kirchsteiger (2004) provide economic models of kindness-based reciprocity as a motivation for sacrificing material payoffs.[6] The kindness of player *i*'s choice with respect to player *j* reflects the highest material payoff available to player *j* given that choice, compared to the range of material payoffs feasible for player *j* after *any* choice by player *i*. In these models, one party considers the degree of kindness implicit in another party's choice and responds favorably to kindness (positive reciprocity) but negatively to unkindness (negative reciprocity). Beliefs about intentions play a central role in the kindness-based models, with psychological game theory serving as the analytic tool (see Geanakoplos, Pearce, and Stacchetti, 1989).

More recently, models have combined kindness-based reciprocity and concern about relative payoffs. Falk and Fischbacher's (forthcoming) model provides a role for intention and uses Fehr and Schmidt's (1999) inequity aversion as the distributional underpinning. Charness and Rabin (2002) present social payoffs as a weighted average of the total material payoffs for the reference group and the minimum material payoff of any member of the group. However, when one party misbehaves with respect to the social standard, the weight assigned to that individual's material payoff diminishes as others withdraw their concern for that party's welfare. The parties also display a taste for punishment, whereby one may sacrifice money to reduce the material payoff of a misbehaving party.[7]

A modest number of experimental studies consider the issues of process and intention. Charness (2004) conducts an experimental labor market in which the attribution for an assigned wage serves as the treatment variable. Workers provide costly effort even when wages are generated by a random mechanism, but a worker's response to a low wage depends on whether a

[6] The appendix to Rabin (1993) includes models that combine distributional and reciprocity concerns.
[7] The model does not include positive reciprocity. The formula is in Appendix B.

self-interested firm intentionally chose the wage, reflecting negative reciprocity.[8] Offerman (2002) studies a game in which a responder can sacrifice money to help or to hurt another party. In one treatment, a first mover intentionally chooses an action either favorable or unfavorable to the responder; in another treatment, this action is chosen through randomization. Offerman finds strong evidence of negative reciprocity and only much weaker evidence of positive reciprocity. However, both the Charness and Offerman studies face the methodological concern mentioned earlier: One party has no choice when there is exogenous determination.

Falk, Fehr, and Fischbacher (2000) test the role of intentions in a design where money may be passed to (and tripled) or taken away from a responder, who may then pass money to or take money from the other player. In one treatment, a self-interested first mover makes the initial choice; in the second treatment, a random draw causes the first choice. This study finds strong support for the role of intention. However, this study employs the within-subject strategy method with multiple options, raising concerns about external validity. A within-subject design may lead to spurious effects when respondents attempt to provide answers to satisfy their perceptions of the experimenter's expectations.[9] Thus, the within-subject strategy method is controversial, with reviewers often split on this form of elicitation.

There are several other experimental studies that yield more equivocal or negative evidence about reciprocity. Bolton, Brandts, and Katok (2000) and Bolton, Brandts, and Ockenfels (1998) find no evidence of positive reciprocity while the latter finds only statistically insignificant evidence of small levels of negative reciprocity. Cox (2000) and Cox and Deck

[8] This paper was the first to explicitly call for a model combining distributional concerns with intention.
[9] Some studies in economics have found substantial differences between with-subject and between-subject designs, as in the survey data in Gneezy (forthcoming) and the experimental data in the working-paper version of Charness, Haruvy, and Sonsino (forthcoming).

(2001) provide mixed evidence regarding positive and negative reciprocity, and find that the results are sensitive to certain aspects of the experimental procedures employed.

A cleaner experimental design should shed useful light on the importance of intentions. We combine an intentional choice with a random factor, thereby improving upon the methodology of comparing responses to choices generated either at random or by a self-interested party; we are unaware of any previous study using this approach. Our design permits comparisons across conditions and identical payoff contexts reached in different manners. In this way, we can control for distributional issues and isolate the role of intention.

Experimental Design

Students at UC Berkeley and UC Santa Barbara served as participants in our experiments. Participants in Treatment 1 were students in an introductory organizational behavior class, and participation was required for course credit. Participants in Treatment 2 were part of the normal experimental population, and we added a show-up fee of $5 to the payoffs earned in the session.

In each session, we placed students on separate sides of a large classroom to begin the experiment. We flipped a coin to determine which side of the room would have the role of *firms* and which side would be *workers*. Each participant then received a packet with the appropriate instructions and decision sheets, each marked with an identifying number used to track decisions. A total of 244 students participated in our eight sessions; each person could participate in only one session.

In our first treatment, each firm received $12 from which to pay a wage to a worker with whom he or she was randomly paired. The firm could choose to pay a low wage ($4) or a high

wage ($8). The wage chosen was subtracted from the $12 endowment and assigned as an interim wage to the worker. The experimental instructions are provided in Appendix A.

We then established the nature of the *business conditions*. We flipped a coin in front of each worker to determine whether these were good or bad. We then either increased or decreased the wage randomly depending on whether business conditions were good or bad, with a 50 percent probability of each condition occurring. All of this was common information.

Under good business conditions, the worker received $2 more than the firm initially chose to pay; under poor business conditions, the worker received $2 less than the firm initially chose to pay. Thus, following the coin toss, there were four possible interim outcomes:

Low wage, poor business conditions	Firm has $8, Worker has $2
Low wage, good business conditions	Firm has $8, Worker has $6
High wage, poor business conditions	Firm has $4, Worker has $6
High wage, good business conditions	Firm has $4, Worker has $10

After seeing the coin flip, the worker chose an effort level: low, medium, or high. Low effort cost the worker $1 and reduced the firm's money by $4. Medium effort cost the worker nothing and left the firm's money unchanged. High effort cost the worker $1 and increased the firm's money by $4.[10] We show all of the resulting possibilities below (all numbers are in $):

[10] We did not inform each firm about either the effort chosen or the outcome of the coin flip, as this seemed to be a cleaner approach in terms of minimizing strategic considerations (e.g., no repeated game considerations) and facilitating the analysis of the observed behavior.

Table 1: Payoffs to the Firm and the Worker in Treatment 1

Wage paid by Firm	Business Conditions	Worker Effort Level		
		Low	Medium	High
Low ($4)	Poor (reduces wages $2)	*4*, **1**	*8*, **2**	*12*, **1**
Low ($4)	Good (raises wages $2)	*4*, **5**	*8*, **6**	*12*, **5**
High ($8)	Poor (reduces wages $2)	*0*, **5**	*4*, **6**	*8*, **5**
High ($8)	Good (raises wages $2)	*0*, **9**	*4*, **10**	*8*, **9**

We performed the experiment twice, matching each firm with a different worker the second time and a giving the firm a new $12 endowment. As per the experimental instructions, we then chose one of these periods at random for actual payment.

A possible concern of this study design is that firm profits differ in the two cases where worker payoffs are the same. Before the worker chooses an effort level, the firm has $8 with low wages and good business conditions, while it has $4 with high wages and poor business conditions. Thus, some of the higher effort we see in good business conditions may be due to a principle of rent-sharing coupled with the firm's poorer condition when it pays high wages.

To test the importance of this possibility, we conducted additional sessions with a slightly different design: The payoff to the firm was $2 more when what we call *wage conditions* were bad (that is, when the firm had to pay the worker $2 less), and the firm paid an additional $2 when these were good. To avoid the possibility of negative earnings, we increased the firm's endowment by $2 to $14. The new base payoffs were:

Low wage, poor wage conditions	Firm has $12, Worker has $2
Low wage, good wage conditions	Firm has $8, Worker has $6
High wage, poor wage conditions	Firm has $8, Worker has $6
High wage, good wage conditions	Firm has $4, Worker has $10

Table 2 shows how the wage conditions and effort levels transformed the base payoffs:

Table 2: Payoffs of the Firm and the Worker in Treatment 2

Wage paid by Firm	Wage Conditions	Worker Effort Level		
		Low	Medium	High
Low ($4)	Poor (reduces wages $2 & raises profits $2)	*8*, **1**	*12*, **2**	*16*, **1**
Low ($4)	Good (raises wages $2 & lowers profits $2)	*4*, **5**	*8*, **6**	*12*, **5**
High ($8)	Poor (reduces wages $2 & raises profits $2)	*4*, **5**	*8*, **6**	*12*, **5**
High ($8)	Good (raises wages $2 & lowers profits $2)	*0*, **9**	*4*, **10**	*8*, **9**

Hypotheses

Our design provides an array of comparisons across conditions, allowing us to examine a number of predictions by various models of social preferences and to test these in formal hypotheses. While the standard model predicts that the worker always chooses to expend medium effort and firms therefore always choose to pay the low wage, few observers of the experimental literature take this prediction very seriously. Our initial null hypotheses are therefore slightly more sophisticated:

> **Hypothesis 1a:** Worker behavior will not differ across the four interim outcomes. In other words, the distribution of choices (punish, don't sacrifice, reward) will not differ significantly for the wage assigned and in different business or wage conditions.
>
> **Hypothesis 1b:** Worker behavior will not differ for high and low assigned wages.

All of the social-preference models discussed earlier predict that these hypotheses will fail. For the distributional models, this anticipated failure results from differences in relative payoffs across these interim outcomes, wherever such differences exist. For the kindness-based reciprocity models, worker behavior should be affected by wages (reflecting intention) but

should be unaffected by the wage or business condition (distributional concerns). Models that combine intention and distribution predict differences for each condition.

Thus, if perceived kindness affects worker behavior but distributional concerns do not, we should expect:

> **Hypothesis 2:** Worker behavior at either level of assigned wage will not differ across the corresponding wage or business conditions.

On the other hand, if only the available choice set matters to the worker, then we should expect no difference in worker behavior between the low-wage, good-condition cell and the high-wage, poor-wage-condition cell. This suggests two hypotheses:

> **Hypothesis 3a:** Worker behavior will not differ across the two identical wage outcomes (low wages and good conditions versus high wages and poor conditions) in either Treatment 1 or Treatment 2.

If workers also care about distribution (but not about the process), then the firm/worker outcome pair is important. In that case, we expect no difference in only the Treatment 2 comparison:

> **Hypothesis 3b:** Worker behavior will not differ across the two identical wage outcomes (low wages and good wage conditions versus high wages and poor wage conditions) in Treatment 2, but may differ in Treatment 1 (low wages and good business conditions versus high wages and poor business conditions).

The Bolton and Ockenfels (2000), Fehr and Schmidt (1999), and Charness and Rabin (2002) models also lead to specific predictions detailed in Appendix B. To summarize, the Bolton and Ockenfels (2000) model predicts medium effort in nearly all cases, except that workers with a strong taste for equality might choose high effort. The Fehr and Schmidt (1999)

model predicts low effort for low wages given sufficient parameter values; there could be high effort for high wages when conditions are good, but medium effort should always be chosen with high wages and poor conditions. The Charness and Rabin (2002) model permits low or medium effort for low wages and high or medium effort for high wages, depending on parameter values; the rate of sacrifice should be the same with low wages regardless of the conditions, while after a high wage we should expect more high effort with good conditions than with poor conditions.

Results

Tables 3 and 4 below display the results for Treatments 1 and 2, respectively. Each Table shows the number of low and high wages for both poor and good conditions, as well the number of effort responses in each category for the wage and conditions. These data are also aggregated across high and low wages, and poor and good conditions.

If Hypothesis 1a were true, we should see punishment and reward rates being the same across all combinations of wage and business conditions. Hypothesis 1b suggests that punishment and reward rates are the same for high and low wages, aggregating across business conditions. Hypothesis 2 predicts that while worker behavior may vary by wage, business conditions should not matter for worker responses to either low or high wages. For Hypotheses 3a and 3b, one should compare the rows in bold; Hypothesis 3a indicates that worker responses in these rows should be the same, while Hypothesis 3b allows for differences due to distributional considerations, as per Bolton and Ockenfels (2000), Fehr and Schmidt (1999), and Charness and Rabin (2002).

Table 3: Treatment 1 Results

Wages	Business Conditions	N	Worker Effort Level % choosing and (n)	
			Punish	Reward
Low	Poor	36	25% (9)	8% (3)
Low	**Good**	**31**	**19% (6)**	**10% (3)**
Low	Total	67	22% (15)	9% (6)
High	**Poor**	**31**	**3% (1)**	**39% (12)**
High	Good	20	0 (0)	60% (12)
High	Total	51	2% (1)	47% (24)
Total	Poor	67	15% (10)	22% (15)
Total	Good	51	12% (6)	29% (15)
Total	All	118	14% (16)	25% (30)

Notes: Medium effort is the omitted category, so % punish + % reward + % medium = 100%.
The two **bold** rows pay identical wages ($6).

The data do not favor either Hypothesis 1a or Hypothesis 1b, as the wage paid by the firm matters substantially. With low wages, workers punish the firm 22 percent of the time and reward the firm 9 percent of the time. These rates are similar in poor business conditions (25 percent punishment; 8 percent reward) and in good business conditions (19 percent punishment; 10 percent reward). In contrast, with high wages, workers punish 2 percent of the time, but reward 47 percent of the time. Again, the rates in poor business conditions (3 percent punishment, 39 percent reward) and good business conditions (no punishment, 60 percent reward) do not differ substantially. Note the disparity between worker behavior aggregated by low and high wages: A Chi-square test of the individual (punish, no sacrifice, reward) tendencies across low and high wages gives $\chi^2(2) = 26.93$, $p < 0.001$. [11]

[11] This test does not take into account that each worker makes two choices. On an individual basis, 13 of the 47 people (28%) who ever faced a low wage sometimes punished a low wage, while only one person (2%) sometimes rewarded a low wage. In contrast, five of the 39 people (13%) who ever faced a high wage sometimes punished a high wage, while 18 people (46%) sometimes rewarded a high wage. No worker ever chose to both punish and

Consistent with Hypothesis 2, good and poor business conditions are not statistically significantly related to effort, although good business conditions result in a bit more high effort and a bit less low effort. Thus, workers in this scenario do not reward managers for workers' good luck. This result is inconsistent with the outcome bias observed in Caplan, Posner, and Cheney (1991) and inconsistent with theories of rent-sharing. We discuss Hypotheses 3a and 3b in more detail later, but note here only that behavior is rather different in the two bold rows.

Given these outcomes, expected firm profits are \$7.47 when the firm pays a low wage and \$5.91 when the firm pays a high wage. It may be that firms paying a high wage make a mistake, in that they do not know the probabilities of reward and retaliation. It could also be the case that some firms prefer either equality or efficiency, considerations that could make a high wage optimal.

The predictions for Treatment 2 are qualitatively the same as for Treatment 1 for Hypotheses 1a, 1b, and 2. However, the payoff design in Treatment 2 eliminates all distributional differences between (Low wage, good condition) and (High wage, poor condition) cells, so that if intentions don't have any affect, worker behavior in these two circumstances should be unambiguously the same. We summarize the results for Treatment 2 in Table 4.

---

reward the same wage in his or her two choices. Re-computing the test using these individual (sometimes punish, never sacrifice, sometimes reward) tendencies, we have $\chi^2(2) = 24.13$, $p < 0.001$.

Table 4: Treatment 2 Results

Wages	Wage Conditions	N	Worker Effort Level % choosing and (n) Punish	Reward
Low	Poor	45	42% (19)	2% (1)
Low	**Good**	**40**	**40% (16)**	**2% (1)**
Low	Total	85	41% (35)	2% (2)
High	**Poor**	**22**	**9% (2)**	**14% (3)**
High	Good	19	5% (1)	58% (11)
High	Total	41	7% (3)	34% (14)
Total	Poor	67	31% (21)	6% (4)
Total	Good	59	29% (17)	20% (12)
Total	All	126	30% (38)	13% (16)

Notes: Medium effort is the omitted category, so % punish + % reward + % medium = 100%. The two **bold** rows pay identical wages.

Once again the wage paid by the firm matters substantially. A Chi-square test of the (punish, don't sacrifice, reward rates) across low and high wages gives $\chi^2(2) = 27.46$, $p < 0.001$.[12] With low wages, workers punish 41 percent of the time and reward 2 percent of the time. These rates are similar with poor wage conditions (42 percent punishment; 2 percent reward) and with good wage conditions (40 percent punishment; 2 percent reward).

In contrast, with high wage conditions, workers punish 7 percent of the time and reward 34 percent of the time. Once again, intention seems to matter, as worker behavior appears different across the bold rows. Compared to Treatment 1, the pattern here is slightly different with respect to conditions: Good and poor wage conditions are not statistically significantly related to effort when the wage is low. However, the rate of reward is much higher when the wage conditions are good than when they are poor ($Z = 3.17$, $p = 0.001$), going against

[12] On an individual basis, 25 of the 47 people (53%) who ever faced a low wage sometimes punished a low wage, while only three people (6%) sometimes rewarded a low wage. In contrast, two of the 37 people (5%) who ever faced a high wage sometimes punished a high wage, while 13 people (35%) sometimes rewarded a high wage. No worker ever chose to both punish and reward a particular wage in her two choices. Re-computing the test using these individual (punish, no sacrifice, reward) tendencies gives $\chi^2(2) = 24.38$, $p < 0.001$.

Hypothesis 2. A major difference between the two treatments is that here good wage conditions mean bad luck for the firm, while in Treatment 1 conditions do not directly affect the firm's payoffs. Thus, when firms pay high wages in this scenario, workers who have had good luck compensate the firms for their bad luck. This result is completely consistent with theories of rent-sharing. On the other hand, negative reciprocity overwhelms this willingness to share when the wage is low; high rates of punishment are nearly identical across wage conditions.

Punishment rates at low wages are higher in Treatment 2 than in Treatment 1; perhaps a low wage does indeed seem less fair in Treatment 2, where the firm's endowment is $14 rather than $12. Reward rates at high wages are similar when wage conditions are good but are lower when wage conditions are poor. It seems plausible that a worker who has just been unlucky may be less willing to reward a firm that has just had good luck. Expected firm profits, given these outcomes, are $8.45 when a low wage is paid and $7.07 when a high wage is paid.

Workers impose higher punishment rates on the firm for paying them low wages, consistent with the story that workers perceive a firm's refusal to pay a wage of $8 as less justifiable when the endowment is $14 than when it is $12. In the latter case, the interim *(Firm,Worker)* payoffs of (4,8) and (8,4) show symmetry, so a social planner would have no preference between these allocations. However, the interim payoff of (6,8) is socially preferred to (10,4), since the distribution is less lopsided. Thus, a firm assigning a low wage in Treatment 2 could seem less fair than a firm doing so in Treatment 1.

Regressions

We supplement our nonparametric analysis by performing regressions regarding wage choice and effort choices. We first consider the determinants of the total wages chosen by each firm in the two periods (recall that the second choice was made without knowledge of the

worker's choice in the first period), with race and gender demographics and some attitudinal measures as explanatory variables.

Table 5: Determinants of total wage
OLS regressions, 3-choice total wage

	(1) Treatment 1	(2) Treatment 1	(3) Treatment 2	(4) Treatment 2
Female	-1.121	-1.413	-0.131	-0.025
	(0.894)	(0.924)	(0.808)	(0.836)
Asian	0.417	0.382	1.502*	0.841
	(1.003)	(1.016)	(0.863)	(0.930)
Other non-white	1.458	1.238	2.142*	1.734
	(1.389)	(1.487)	(1.156)	(1.175)
Political views		0.393		0.626**
		(0.404)		(0.310)
Like market society		-0.088		0.346*
		(0.304)		(0.212)
Should help weaker people		-0.551*		-0.183
		(0.338)		(0.300)
Constant	11.734***	12.320***	9.937***	7.272***
	(0.941)	(1.964)	(0.770)	(1.755)
Observations	59	59	63	63
R^2	0.044	0.095	0.075	0.167

Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
Political views (conservative-liberal), Like market society (agree-disagree), and should Help weaker people (agree-disagree) are on 1-7 scales. Low effort = 0, Medium effort = 1, and High effort = 2

It is formally correct to use an ordered probit to take into account that responses are not cardinal; results were very similar when we used that specification and so here and elsewhere we present the easier-to-interpret linear specifications.

In Treatment 1, there are only modest and insignificant effects from the gender and demographic variables, with females offering slightly lower wages and non-whites offering slightly higher wages than the baseline white group. The more a firm agrees that society should help its weaker members, the higher the wage offered; firms that strongly agree offer wages 42% higher than firms that strongly disagrees, and this effect is marginally significant. Similarly,

firms with the most liberal political views offer wages 19% higher than firms with the most conservative political views, but this effect is not statistically significant.

In Treatment 2, non-whites offer slightly higher wages, and this effect is marginally significant in regression (3); however, this statistical significance vanishes in regression (4), when attitudinal variables are included. We again see some effect from the attitudinal variables, although here political views and feelings about living in a market society are significant, while beliefs about helping weaker members of society are not. Firms with the most liberal political views offer wages 48% higher than firms with the most conservative political views; firms that least approves of a market society offer wages 28% higher than firms that most approve.

We also consider the determinants of the effort chosen by each worker, using effort choice as the dependent variable and business/wage condition dummies, race and gender demographics, and attitudinal measures as explanatory variables.

Table 6: Determinants of effort
Random-effects regressions, 3-choice effort

	(1) Treatment 1	(2) Treatment 1	(3) Treatment 2	(4) Treatment 2
High wage	0.465***	0.502***	0.386***	0.389***
	(0.123)	(0.130)	(0.130)	(0.133)
Good conditions	0.067	0.066	-0.015	0.005
	(0.117)	(0.125)	(0.109)	(0.113)
High wage*Good conditions	0.168	0.156	0.544***	0.544***
	(0.193)	(0.202)	(0.189)	(0.192)
Female		0.060		-0.143
		(0.164)		(0.129)
Asian		-0.254*		0.132
		(0.147)		(0.134)
Other non-white		-0.204		0.187
		(0.177)		0.160
Political views		0.006		-0.065
		(0.055)		(0.043)
Like market society		-0.033		0.025
		(0.048)		(0.034)
Help weaker people		0.001		-0.039
		(0.039)		(0.043)
Constant	0.860***	1.049***	0.625***	0.822***
	(0.090)	(0.313)	(0.083)	(0.236)
Observations	118	112	126	124
Number of Groups	59	56	63	62
Overall R^2	0.242	0.313	0.284	0.356

Standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%
Political views (conservative-liberal), Like market society (agree-disagree), and should Help weaker people (agree-disagree) are on 1-7 scales. Low effort = 0, Medium effort = 1, and High effort = 2

These regressions are generalized least squares regressions reflecting the three feasible effort choices; we use robust standard errors to account for the fact that each worker makes two effort choices.

In both treatments and in all specifications, a high wage induces an average effort level considerably higher than does a low wage, and this difference is highly significant. Confirming the pattern seen in Table 1, business conditions have no serious effect on effort provision in Treatment 1, with either a high wage or a low wage. On the other hand, while wage conditions

do not affect the effort response to a low wage in Treatment 2, the interaction term is highly significant in regressions (3) and (4). Thus, we can confirm that effort is sensitive to wage conditions when the worker is assigned a high wage, but not otherwise. None of the demographic or gender variables have an effect, with the possible exception of the Asian dummy variable. Asians choose somewhat lower effort in Treatment 2 than the baseline; this contrasts with the findings for the business administration class in Treatment 1, where Asians provide slightly higher effort than the baseline.

Discussion

Hypotheses 3a and 3b represent the centerpiece of our experimental design: Does the path leading to a specific worker's wage influence the worker's behavior? If only the worker's payoff matters, in both Treatments 1 and 2, we should expect the same distribution of worker choices among those who receive low wages in good conditions as we find among those who receive high wages in poor conditions. If we find that a different distribution of worker choices in Treatment 1 but not in Treatment 2, this would indicate that distributional considerations play a role, but that intention is unimportant.

We see that in both Treatment 1 and Treatment 2, worker behavior is considerably different across these combinations, rejecting Hypotheses 3a and 3b. Chi-square tests for whether the distribution of effort choices is the same give $\chi^2(2) = 9.37$, $p = 0.009$ and $\chi^2(2) = 6.48$, $p = 0.039$, for the respective tests.

Effort depends strongly on the wage received in all cases. In Treatment 1, the business conditions, which affect only the workers' payoffs, do not appreciably affect effort choices. In Treatment 2, on the other hand, a worker's bad luck is a firm's good luck. The matter of luck doesn't seem to affect the urge to punish that is engendered by a low wage; however, there is

greater reluctance to reward a firm for offering a high wage when the wage conditions are poor, so that the firm has already been 'rewarded' by good luck at the worker's expense.

We are pleased to observe strong effects for intention in both the standard university subject pool and students from a class, suggesting at least some degree of robustness for our results. It appears that firms were more generous in Treatment 1 and workers were less likely to reward and more likely to punish in Treatment 2. One might have expected ex ante that a high wage was more frequent in Treatment 2, where the firm's endowment is $2 larger. The greater proportion of high wages in Treatment 1 (43% vs. 33%) perhaps reflects the difference in how the experiment was perceived by students partially motivated by a class requirement and by students responding to a notice for a paid experiment, although the fact that conditions affect the firms' earnings in Treatment 2 might have also helped to induce the lower wages there.

Let us consider the predictions of the fairness mentioned earlier. Workers do not tend to reward firms that pay low wages, in accordance with all the models. The substantial rate of low worker effort among those receiving low wages in poor conditions in both treatments is consistent with all the models except for Bolton and Ockenfels (2000). Punishment rates in the low wage cases with poor versus good wage conditions differ very little in the two treatments. This is consistent with the kindness-based reciprocity models but not with the distributional models. The motivation for low effort seems to stem from considerations of negative intention rather than from unfavorable relative payoffs per se.

With respect to reward behavior among workers receiving high wages in poor wage conditions, the Fehr and Schmidt (1999) model predicts that the rate of reward in Treatment 1 should be smaller than the punishment rate among those receiving low wages in both good and

poor wage conditions.13 However, the 39 percent reward rate exceeds the 25 percent and 19 percent punishment rates among those receiving low wages in poor or good wage conditions, respectively.

The reward rate does appear to be sensitive to the wage conditions, particularly in Treatment 2; the role of relative material payoffs is not consistent with Rabin (1993) and Dufwenberg and Kirchsteiger (2004). In Treatment 2, Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) predict that no workers receiving high wages in poor wage conditions will choose to expend high effort, and in our experiment only 10 percent choose to reward the firm with high effort. Distributional considerations appear to have more weight when it comes to the choice to provide high effort. Consistent with the Charness and Rabin (2002) model, the rate of sacrifice for punishment is unaffected by business or wage conditions, while we see substantially more reward with good conditions than with poor conditions.

Conclusion

We find that workers frequently reward and almost never punish firms if they perceive good intentions, even when their own outcomes are not particularly good. Workers frequently punish and almost never reward if they perceive bad intentions, even when their own outcomes are not particularly bad.

What is novel in our experiment is that an identical material payoff for workers can be received due to different combinations of intention and luck. This design permits us to cleanly compare the effects of material payoffs and intentions. To test for the effects of distributional concerns, we can compare workers' behavior under different business or wage conditions at each

[13] This reflects the cut-off values for punishment or reward, as shown in Appendix B. Specifically, if $2\alpha - 3\beta > 1$, then $2\alpha - \beta$ must also be greater than 1; if $2\alpha - 3\beta > 1$, then α must be greater than 1/3.

wage level. We find that effort after a low wage does not depend much on distributional considerations. However, employee behavior depends strongly upon the path that leads to the interim outcome. That is, in contrast to most economic theories, intentions matter more than do material payoffs. In contrast to some important recent fairness theories (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), intentions matter more than relative payoffs.

These results are thus broadly consistent with models of kindness-based reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). Nevertheless, the pure reciprocity models do not predict that reward behavior in response to a high wage will be sensitive to external conditions. Some degree of rent-sharing exists, particularly in Treatment 2, where favorable wage conditions represent a loss for the firm. Only models that combine distributional concerns with reciprocity motivations can account for the observed patterns of effort choices.

Our findings on intentions' importance are consistent with the broader literature on procedural justice (Lind and Tyler, 1988). That is, most people like a system that will on average deliver good outcomes (at least to those who deserve them), even if the outcome is not good in each specific case. A manager who pays a high wage from his or her own pocket clearly tries to deliver a good outcome for the worker; consistent with the literature on procedural justice, we find workers frequently reciprocate in kind.

Our results also emphasize how *impression management* can affect employees' and customers' responses to bad outcomes (Brockner et al., 1994). The managerial implication is clear: If outcomes are poor, be sure that you can justify to employees or other stakeholders (such as investors and customers) that the firm tried to do the right thing.

The patterns we observe are consistent with recent experimental results by Falk, Fehr, and Fischbacher (2000), where respondents replied to 13 contingent options. Because revenge is

often motivated by visceral reactions that are hard to simulate when answering contingent questions, our study design enhances confidence in the importance of revenge. In fact, in some but not all studies (e.g., Brandts and Charness, 2003, but not Brandts and Charness, 2000), punishment rates are much higher when people answer directly, not contingently.[14]

Limitations and Possible Extensions. A number of this study's features help highlight the role of intentions. Yet, intentions may have quite different effects in different settings.

In this experiment, intended high wages' failure to translate into actual high wages results explicitly from luck, not effort—the students saw the experimenter flip a coin. People may be less forgiving of failed good intentions if they feel the one with good intentions had low levels of effort or skill. This effect seems particularly likely in situations where norms of reciprocity hold. To take an extreme case, assume Bob does Alice a favor but does it poorly. At the same time, assume that the favor creates an expectation that in the future, Alice will do a costly favor for Bob. In this extreme case, Bob's initial favor, which might have involved the best of intentions, makes Alice worse off.

The explicit coin flip in our study also makes it easier for workers to assess matters. In the field, people may have a difficult time discerning intentions, given that they often see only outcomes. An important extension of this research examines repeated interactions where people do not see intentions directly but can identify patterns of outcomes consistent with good or bad intentions.

This signal extraction problem is exacerbated by "hindsight bias" or "outcome bias": People who observe a poor outcome often assume that the other party is at fault. For example, in one study, doctors reviewed cases with the same facts but with randomly assigned outcomes.

[14] See Hsee, Blount, Loewenstein, and Bazerman (1999) for a review of within- versus between-subject designs.

Doctors who rated cases with randomly assigned bad outcomes tended to rate the care as substandard much more than did doctors who rated cases they believed to have a neutral outcome (Caplan, Posner, and Cheney, 1991). It is plausible, though not yet shown, that people have an outcome bias for estimating intentions as well as care.

Finally, relative payoffs do not have an important role in this study. For example, workers do not consistently share the benefits of good luck with firms. However, good luck may be shared more often in ongoing relationships than in brief lab studies such as this.

References

Adams, J. (1965). "Inequity in Social Exchange." In L. Berkowitz (ed.), *Advances in Experimental Social Psychology,* Vol. 2, 267-299. New York: Academic Press.

Andreoni, J., and J. Miller. (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, **70**, 737-753.

Bertrand, M., and S. Mullainathan. (1999). "Is There Discretion in Wage Setting? A Test Using Takeover Legislation." *Rand Journal of Economics*, **30**, 535-554.

Bewley, T. (1999). *Why Wages Don't Fall During a Recession.* Cambridge: Harvard University Press.

Blanchflower, D., A. Oswald, and M. Garrett. (1990). "Insider Power in Wage Determination." *Economica*, **57**, 143-170.

Blount, S. (1995). "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes*, **63**, 131-144.

Bolton, G., J. Brandts, and E. Katok. (2000). "How Strategy Sensitive Are Contributions? A Test of Six Hypotheses in a Two-Person Dilemma Game." *Economic Theory*, **15**, 367-387.

Bolton, G., J. Brandts, and A. Ockenfels. (1998). "Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game." *Experimental Economics*, **1,** 207-219.

Bolton, G., and A. Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, **90**, 166-193.

Brandts, J., and G. Charness. 2000. "Hot vs. Cold: Sequential Responses in Simple Experimental Games." *Experimental Economics*, **2**, 227-238.

------. (2003). "Truth or Consequences: An Experiment." *Management Science*, **49**, 116-130.

Brockner J., M. Konovsky, R. Cooper-Schneider, R. Folger, C. Martin, and R. Bies. (1994). "Interactive Effects of Procedural Justice and Outcome Negativity on Victims and Survivors of Job Loss." *Academy of Management Journal*, **37**, 397-409.

Caplan, R., K. Posner, and F. Cheney. (1991). "Effect of Outcome on Physician Judgements of Appropriateness of Care." *Journal of the American Medical Association,* **265**, 1957-1960.

Charness, G. (2000). "Responsibility and Effort in an Experimental Labor Market." *Journal of Economic Behavior and Organization*, **42**, 375-384.

------. (2004). "Attribution and Reciprocity in an Experimental Labor Market." *Journal of Labor Economics*, **22**, 626-647.

Charness, G., and B. Grosskopf. (2001). "Relative Payoffs and Happiness: An Experimental Study." *Journal of Economic Behavior and Organization*, **45**, 301-328.

Charness, G, E. Haruvy, and D. Sonsino (forthcoming),"Social Distance and Reciprocity: The Internet vs. the Laboratory" *Journal of Economic Behavior and Organization*.

Charness, G., and D. Levine. (2002). "Changes in the Fair Employment Contract: Evidence from a Quasi-Experiment." *Journal of Economic Behavior and Organization*, **47**, 391-405.

Charness, G., and M. Rabin. (2002). "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, **117**, 817-869.

Cox, J. (2000). "Trust and Reciprocity: Implications of Game Triads and Social Contexts." Mimeo.

Cox, J., and C. Deck. (2001). "On the Nature of Reciprocal Motives." Mimeo.

Dufwenberg, M., and G. Kirchsteiger. (2004). "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, **47**, 268-298.

Falk, A., E. Fehr, and U. Fischbacher. (2000). "Testing Theories of Fairness -- Intentions Matter." Working Paper No. 63, University of Zürich.

Falk, A., and U. Fischbacher. (forthcoming). "A Theory of Reciprocity." *Games and Economic Behavior*.

Fehr, E., and S. Gächter. (2000). "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, **14**, 159-181.

Fehr, E., and K. Schmidt. (1999). "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, **114**, 817-868.

Frohlich, N., and J. Oppenheimer. (1984). "Beyond Economic Man: Altruism, Egalitarianism, and Difference Maximizing." *Journal of Conflict Resolution*, **28**, 3-24.

-----. (1992). *Choosing Justice.* Berkeley: University of California Press.

Geanakoplos, J., D. Pearce, and E. Stacchetti. (1989). "Psychological Games." *Games and Economic Behavior*, **1**, 60-79.

Gneezy, U. (forthcoming), "Deception: The Role of Consequences," *American Economic Review*.

Goranson, R., and L. Berkowitz. (1966). Reciprocity and Responsibility Reactions to Prior Help. *Journal of Personality and Social Psychology*, **3**, 227-232.

Gouldner, A. (1960). "The Norm of Reciprocity. *American Sociological Review*, **25**, 161-178.

Greenberg, M., and D. Frisch. (1972). "Effect of Intentionality on Willingness to Return a Favor." *Journal of Experimental Social Psychology*, **8**, 99-111.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Hsee, C., S. Blount, G. Loewenstein, and M. Bazerman. (1999). "Preference Reversals Between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis." *Psychological Bulletin*, **125**, 576-590.

Huang, P. (2000). "Reasons within Passions: Emotions and Intentions in Property Rights Bargaining." *Oregon Law Review*, **79**, 435-478.

Kahn, A. and T. Tice. (1973). "Returning a Favor and Retaliating Harm: The Effects of Stated Intentions and Actual Behavior." *Journal of Experimental Social Psychology*, **9**, 43-56.

Kahneman, D., J. Knetsch, and R. Thaler. (1986). "Fairness as a Constraint on Profit-Seeking: Entitlements in the Market." *American Economic Review*, **76**, 728-741.

Kelley, H. (1972). "Attribution in Social Interaction." In Jones et al. (ed.), *Attribution: Perceiving the Causes of Behavior*, 1-26. Morristown, NJ: General Learning Press.

Kelley, H., and A. Stahelski. (1970). "Errors in Perception of Intentions in a Mixed-Motive Game." *Journal of Experimental Social Psychology*, **6**, 379-400.

Kritikos, A., and F. Bolle. (1999). "Approaching Fair Behavior: Self-Centered Inequality Aversion Versus Reciprocity and Altruism." Mimeo.

Levine, D. (1993). "What Do Wages Buy?" *Administrative Science Quarterly*, **38**, 462-483.

Lind, E., and T. Tyler. (1988). *The Social Psychology of Procedural Justice*. New York: Plenum Press.

Morgenstern, A. (2004), "Efficiency concerns and incentive provision - an experimental study," *Economics Letters*, 83, 335-341.

Nisbett, R., and L. Ross. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Offerman, T. (2002). "Hurting Hurts More Than Helping Helps." *European Economic Review*, **46**, 1423-1437.

Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, **83**, 1281-1302.

Ross, L. (1977). "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process." In L. Berkowitz (ed.), *Advances in Experimental Social Psychology,* Vol. 10, 174-221. New York: Academic Press.

Thibaut, J., and H. Riecken. (1955). "Some Determinants and Consequences of the Perception of Social Causality." *Journal of Personality*, **24**, 113-133.

Tyler, T. (1988), "What is Procedural Justice? Criteria used by Citizens to Assess the Fairness of Legal Procedures," Law Society Review, **22**, 301-355.

Waldfogel, J. (1993). "The Deadweight Loss of Christmas." *American Economic Review*, **83**, 1328-1336.

------. (1996). "The Deadweight Loss of Christmas: Reply." *American Economic Review*, **86**, 1306-1308.

APPENDIX A: Instructions

Instructions for The Firm (Treatment 1)

You are a *firm* and have been endowed with $12 from which you pay a wage to a *worker* with whom you are paired. You may choose to pay a low wage ($4) or a high wage ($8). Whichever wage you choose will be subtracted from your $12 endowment.

The wage you choose is paid to the worker with whom you are paired. However, this wage will randomly be either increased or decreased depending on whether company sales are high or low, with a 50% probability of each condition occurring.

If business conditions are good, then the worker will receive $2 more than you have initially paid, while if business conditions are poor the worker will receive $2 less than you have initially paid. Both good and poor business conditions are equally likely, depending on the flip of a fair coin. Thus, there are four possible outcomes after a coin is flipped to determine whether business conditions are good or poor:

Low wage, poor business conditions	Firm has $8, Worker has $2
Low wage, good business conditions	Firm has $8, Worker has $6
High wage, poor business conditions	Firm has $4, Worker has $6
High wage, good business conditions	Firm has $4, Worker has $10

Next, the worker chooses an effort level: low, medium, or high. Low effort costs the worker $1, and reduces the firm's money by $4. Medium effort costs the worker nothing and leaves the firm's money unchanged. High effort costs the worker $1, and increases the firm's money by $4. All of the resulting possibilities are shown below:

Low wage, poor business conditions		
Effort	*Firm*	*Worker*
Low	4	1
Medium	8	2
High	12	1

Low wage, good business conditions		
Effort	*Firm*	*Worker*
Low	4	5
Medium	8	6
High	12	5

High wage, poor business conditions		
Effort	*Firm*	*Worker*
Low	0	5
Medium	4	6
High	8	5

High wage, good business conditions		
Effort	*Firm*	*Worker*
Low	0	9
Medium	4	10
High	8	9

Procedure:

You choose the wage for the worker with whom you are paired. A coin flip (separate for each worker) then determines whether the business conditions are good or poor. The worker will know the result of this coin flip and the wage you have chosen at the time of his or her choice of effort. The combination of the wage you choose and the effort chosen by the worker will determine the outcome.

We will do this twice, with each firm matched with a different worker the 2nd time, with a new $12 endowment for the firm. We will then choose one of these periods at random for actually payment. You will be paid individually and privately. Thank you for your participation.

Instructions for the Worker (Treatment 1)

You are a *worker* and are paired with a *firm* that has been endowed with $12 from which to pay you a wage. The firm may choose to pay a low wage ($4) or a high wage ($8). Whichever wage is chosen is subtracted from the firm's $12 endowment.

The amount you receive will be higher or lower than the amount the firm paid, depending on business conditions. Half the time (determined by flipping a fair coin) you will receive $2 more than the firm paid, and half the time you will receive $2 less. Thus, there are four possible outcomes:

Low wage, poor business conditions	Worker has $2, Firm has $8
Low wage, good business conditions	Worker has $6, Firm has $8
High wage, poor business conditions	Worker has $6, Firm has $4
High wage, good business conditions	Worker has $10, Firm has $4

Next, you choose an effort level: low, medium, or high. Low effort costs $1, and reduces the firm's money by $4. Medium effort costs nothing and leaves the firm's money unchanged. High effort costs $1, and increases the firm's money by $4. All of the resulting possibilities are shown below:

Low wage, poor business conditions		
Effort	*Worker*	*Firm*
Low	1	4
Medium	2	8
High	1	12

Low wage, good business conditions		
Effort	*Worker*	*Firm*
Low	5	4
Medium	6	8
High	5	12

High wage, poor business conditions		
Effort	*Worker*	*Firm*
Low	5	0
Medium	6	4
High	5	8

High wage, good business conditions		
Effort	*Worker*	*Firm*
Low	9	0
Medium	10	4
High	9	8

Procedure:

The firm with whom you are paired chooses your wage. A coin flip (separate for each worker) then determines your business conditions. You will know the result of this coin flip and the wage chosen at the time. The combination of the wage chosen by the firm and the effort you choose will determine the outcome.

We will do this twice, with each firm matched with a different worker the 2nd time, with a new $12 endowment for the firm. We will then choose one of these periods at random for actually payment. You will be paid individually and privately. Thank you for your participation.

Instructions for the Firm (Treatment 2)

You are a *firm* and have been endowed with $14 from which you pay a wage to a *worker* with whom you are paired. You may choose to pay a low wage ($4) or a high wage ($8). Whichever wage you choose will be subtracted from your $14 endowment.

The wage you choose will randomly be either increased or decreased depending on external conditions, with a 50% probability of each condition occurring.

If wage conditions are good, then the wage you actually pay the worker will be $2 more than you had initially chosen, while if wage conditions are poor then the wage you actually pay the worker will be $2 less than you had initially chosen. Both good and poor wage conditions are equally likely, depending on the flip of a fair coin. Thus, there are four possible outcomes after a coin is flipped to determine whether wage conditions are good or poor:

Low wage, poor wage conditions	Firm has $12, Worker has $2
Low wage, good wage conditions	Firm has $8, Worker has $6
High wage, poor wage conditions	Firm has $8, Worker has $6
High wage, good wage conditions	Firm has $4, Worker has $10

Next, the worker chooses an effort level: low, medium, or high. Low effort costs the worker $1, and reduces the firm's money by $4. Medium effort costs the worker nothing and leaves the firm's money unchanged. High effort costs the worker $1, and increases the firm's money by $4. All of the resulting possibilities are shown below:

Low wage, poor wage conditions		
Effort	*Firm*	*Worker*
Low	8	1
Medium	12	2
High	16	1

Low wage, good wage conditions		
Effort	*Firm*	*Worker*
Low	4	5
Medium	8	6
High	12	5

High wage, poor wage conditions		
Effort	*Firm*	*Worker*
Low	4	5
Medium	8	6
High	12	5

High wage, good wage conditions		
Effort	*Firm*	*Worker*
Low	0	9
Medium	4	10
High	8	9

Procedure:

You choose the wage for the worker with whom you are paired. A coin flip (separate for each worker) then determines whether the wage conditions are good or poor. The worker will know the result of this coin flip and the wage you have initially chosen at the time of his or her

choice of effort. The combination of the wage paid and the effort chosen by the worker will determine the outcome.

We will do this twice, with each firm matched with a different worker the 2nd time, with a new $14 endowment for the firm. We will then choose one of these periods at random for actually payment. You will be paid individually and privately. Thank you for your participation.

Instructions for the Worker (Treatment 1)

You are a *worker* and are paired with a *firm* that has been endowed with $14 from which to pay you a wage. The firm may choose to pay a low wage ($4) or a high wage ($8). Whichever wage is chosen is subtracted from the firm's $14 endowment.

The amount actually paid will be higher or lower than the amount the firm paid, depending on wage conditions. Half the time (determined by flipping a fair coin) the wage paid will be $2 more than the firm chose initially, and half the time the wage paid will be $2 less. Thus, there are four possible outcomes:

Low wage, poor wage conditions	Worker has $2, Firm has $12
Low wage, good wage conditions	Worker has $6, Firm has $8
High wage, poor wage conditions	Worker has $6, Firm has $8
High wage, good wage conditions	Worker has $10, Firm has $4

Next, you choose an effort level: low, medium, or high. Low effort costs $1, and reduces the firm's money by $4. Medium effort costs nothing and leaves the firm's money unchanged. High effort costs $1, and increases the firm's money by $4. All of the resulting possibilities are shown below:

Low wage, poor wage conditions		
Effort	*Worker*	*Firm*
Low	1	8
Medium	2	12
High	1	16

Low wage, good wage conditions		
Effort	*Worker*	*Firm*
Low	5	4
Medium	6	8
High	5	12

High wage, poor wage conditions		
Effort	*Worker*	*Firm*
Low	5	4
Medium	6	8
High	5	12

High wage, good wage conditions		
Effort	*Worker*	*Firm*
Low	9	0
Medium	10	4
High	9	8

Procedure:

The firm with whom you are paired chooses your wage. A coin flip (separate for each worker) then determines your wage conditions. You will know the result of this coin flip and the wage initially chosen at the time of your choice of effort. The combination of the wage paid and the effort you choose will determine the outcome.

We will do this twice, with each firm matched with a different worker the 2nd time, with a new $14 endowment for the firm. We will then choose one of these periods at random for actually payment. You will be paid individually and privately. Thank you for your participation.

APPENDIX B: Models of Utility and Their Predictions

We present the two-person version of some economic models of utility, along with their predictions. As a benchmark, the standard neoclassical model predicts medium effort and low wages in all cases.

Bolton and Ockenfels (2000):

Bolton and Ockenfels assume people prefer equality, so person *i*'s utility can be expressed as a function:

$$U_i \equiv v(\pi_i, \sigma_i),$$

where π_i is player *i*'s payoff, and $\sigma_i = \pi_i/(\pi_i + \pi_j)$ is player *i*'s share of the total payoff to the two players. They assume that all players believe that even shares are fair, so with two players and positive payoffs, utility declines as $|\sigma_i - 1/2|$ increases. The following assumptions are made:

$$v_{i1}(\pi_i, \sigma_i) \geq 0,\ v_{i11} \leq 0,\ v_{i2} = 0 \text{ for } \sigma_i = 1/2, \text{ and } v_{i22}(\pi_i, \sigma_i) < 0.$$

The model's predictions for effort in our context:

Treatment 1:

Low, Poor	All medium effort
Low, Good	Small possibility of low effort, otherwise all medium effort
High, Poor	All medium effort
High, Good	High effort from workers with strong taste for equality; otherwise medium effort

Treatment 2:

Low, Poor	All medium effort
Low, Good	Small possibility of low effort, otherwise all medium effort
High, Poor	Same as for Low, Good
High, Good	High effort from workers with strong taste for equality; otherwise medium effort

Fehr and Schmidt (1999):

Fehr and Schmidt's model is similar in spirit to that of Bolton and Ockenfels, but assumes that people find being paid less than their fair share is at least as painful as being overpaid:

$$U_i \equiv \pi_i - \alpha_i(\max\{\pi_j - \pi_i\},0) - \beta_i(\max\{\pi_j - \pi_i\},0).$$

Here α_i reflects the extent to which *i* dislikes being behind, and β_i reflects the extent to which *i* dislikes being ahead. They assume that $\alpha_i \geq \beta_i$ (being behind is at least as painful as being ahead) and $\beta_i < 1$ (people do not burn their own money just to reach equality).

The model's predictions for effort in our context:

Treatment 1:

Low, Poor	Low effort if $\alpha > 1/3$, otherwise medium effort
Low, Good	Low effort if $2\alpha - \beta > 1$, otherwise medium effort
High, Poor	All medium effort ($3\alpha > 2\beta$, since $\alpha > \beta$)
High, Good	High effort if $5\beta > 1$, otherwise medium effort

Treatment 2:

Low, Poor	Low effort if $\alpha > 1/3$, otherwise medium effort
Low, Good	Low effort if $2\alpha - \beta > 1$, otherwise medium effort
High, Poor	All medium effort ($7\alpha > 2\beta$, since $\alpha > \beta$)
High, Good	High effort if $5\beta > 1$, otherwise medium effort

Charness and Rabin (2002):

The Charness and Rabin model includes most of the features of the previous models and also includes a concern for the other party's intentions, as measured by d_j. This variable represents the other player's level of demerits, where the higher the value of d_j, the less *i* thinks the other player deserves. Demerits, d_j, are allocated when people act selfishly at the expense of efficiency and equity. The utility function that takes into account demerits is:

$$U_i \equiv (1-\lambda)\cdot\pi_i + \lambda\cdot[\delta \min[\pi_i, \pi_j + b\cdot d_j] + (1-\delta)\cdot(\pi_i + \max[1-k\cdot d_j, 0]\pi_j) - f\cdot d_j\cdot\pi_j].$$

Utility is a function of own material payoffs (with weight $1-\lambda$) and social payoffs (with weight λ). Social payoffs, in turn, are an average of concern for the lowest payoff (and in a two-player game, this parameter also captures equality) with weight δ and concern for efficiency with weight $1-\delta$. The weights λ and δ range between 0 and 1, inclusive. The non-negative parameter *f* captures a player's taste for punishing miscreants. The non-negative parameters *b* and *k* capture the notion that low payments to miscreants do not provide as much disutility as low payments to well-behaving others.

The model's predictions for effort in our context:

Treatment 1:

Low, Poor	Low effort if *f* is high, otherwise medium effort
Low, Good	Same as Low, Poor
High, Poor	High effort if $\lambda(3 - 2\delta) > 1$, otherwise medium effort
High, Good	High effort if $\delta(3 + \lambda) > 1$, otherwise medium effort

Treatment 2:

Low, Poor	Low effort if *f* is high, otherwise medium effort
Low, Good	Same as Low, Poor
High, Poor	High effort if $\lambda(3 - 4\delta) > 1$, otherwise medium effort
High, Good	High effort if $\delta(3 + \lambda) > 1$, otherwise medium effort