INTENTION AND STOCHASTIC OUTCOMES: AN EXPERIMENTAL STUDY

Gary Charness and David I. Levine*

Do people care about intentions – even when good intentions do not produce good results? In our experiments we find that rates of punishment and reward react strongly to intentions (the wage a firm decides to pay) and more modestly to distributional outcomes (the higher or lower wage actually received including the stochastic component). For example, workers who end up receiving medium wages respond much more positively when this resulted from the firm offering a high wage but bad luck lowered the worker's pay than when this resulted from the firm offering a low wage and good luck raised the pay.

Outcomes for any given situation often depend on a combination of intentional choices and luck. To assess the importance of good intentions, 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.[1]

From Adam's (1965) classic equity theory to recent economic models of fairness, many social scientists have extended the assumption of self-interested preferences to include the idea that judgements and actions reflecting fairness are based on relative outcomes; see Fehr and Schmidt (1999); Bolton and Ockenfels (2000). However, while relative outcomes often play a role, the intentions that led to an outcome often matter as well.

There is a strong case that intentions matter: for example, intentions play a major role in US legal codes, often determining whether punitive damages are to be awarded in civil cases and the degree of punishment in criminal cases (Huang, 2000). For example, manslaughter receives a lighter sentence than second-degree murder, which in turn receives a lighter sentence than first-degree murder. Each crime results in a deceased victim, but they differ in the jury's perceptions of the perpetrator's intentions.

Kahneman *et al.* (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 or whether the action was clearly motivated by selfish intentions. Bewley notes that because morale is important for workplace performance, 'reciprocity and even humanitarian feelings have an impact on behaviour' (1999, p. 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.

Economic models of kindness-based reciprocity offer perceived kindness or unkindness as the primary motivation for why people choose not to maximise their own material payoffs; see Rabin (1993); Dufwenberg and Kirchsteiger (2004). The more recent models proposed by Charness and Rabin (2002) and Falk and Fischbacher (2006) enrich the outcome-based models by incorporating both perceived intentions and distributional concerns and reactions.

We provide a direct test of intention's influence on behaviour by explicitly separating intentions from outcomes. Our study includes a stochastic component in determining wages 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.

Our design permits us to separate the influences of intentions and distribution without various confounding factors. In our context, the first-mover's intention is clear; to test for the effects of distributional concerns, we can compare workers' behaviour after good or bad luck at each wage level. One novel element in our experiment is that prior to deciding how to respond, a worker (and in one treatment, also the firm) can receive an identical material payoff due to different combinations of intention and luck. Models of utility that include only distributional concerns (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) can possibly explain workers' sacrifice here but could not explain worker behaviour being affected by the path taken to arrive at this payoff.

Our methodology differs from methods in previous studies (discussed in detail in the next Section). While previous studies use several designs, most have considered foregone alternatives as indicators of intention. For example, consider this pair of games from Charness and Rabin (2002): in one, player B unilaterally chooses between (A, B) payoffs of (750, 375) or (400, 400). Nearly half of all Bs choose the first choice; that is, to sacrifice 25, thereby adding 350 to A's payoff. In the second game, A can choose an outside option giving (A, B) payoffs of (750, 0), or A can elect to pass the choice to B. Bs in this game choose between (750, 375) and (400, 400), as in the first game. If one believes that good intentions matter, foregoing the outside option might seem to be a kind act, as doing so guarantees B a non-zero payoff. Thus, 'positive reciprocity' should lead to a higher rate of sacrifice by Bs in the second game. However, the rate of sacrifice by Bs is actually less, about 38%. On the basis of this and similar results, Charness and Rabin (2002) puts forward a model that does not include positive reciprocity.

As it turns out, there may be some confounding factors, and perhaps positive reciprocity is indeed present. In the second game, B presumably forms a belief about why A chose to forego the outside option. Possibly B believes A is being kind; however, a self-serving rationalisation might be that A has foregone this option because he or she really prefers equal payoffs, so B should choose (400, 400). A second concern is that B may feel that A had a chance to receive a good payoff and chose not to take it, so that one is

relieved of some of the responsibility of being 'fair'. Thus, there may be a tendency for Bs to sacrifice more in the first game, where A is helpless.[2] We avoid these issues by featuring a direct test whereby two different routes lead to the same interim outcome,[3] so that we may observe how responses vary across intention without requiring indirect inferences or denying one party the opportunity to make a choice.

We find that the rates of punishment and reward are highly sensitive to the wage selected by the employer. More interestingly, despite identical wages (and even identical relative payoffs) responders' behaviours differ strongly. In our data, intention typically matters much more than distributional considerations do.

1. Background and Previous Work

Economists have used two general approaches to study intentions experimentally, with some methodological variations within approaches.[4]

The first approach used involves comparing responses to a decision made intentionally by a self-interested party to a 'decision' made by a random draw from a distribution. For example, gift-exchange results typically show a strong and consistent positive relationship between the wage offered by an experimental 'firm' and the costly 'effort' provided by the paired 'worker'. However, it is possible that this pattern is driven not by some form of reciprocity, but instead by a desire to share the wealth inexpensively.[5] If intention matters, we should expect a different pattern of responses depending on how the wage was chosen. Charness (1996, 2004) finds that workers indeed respond less favourably to low wages chosen by self-interested parties than to the same low wages generated by a random draw (negative reciprocity); however, no evidence is found of a difference in responses to higher wages (positive reciprocity).[6]

Offerman (2002) permits responders to sacrifice money to help (hurt) a party after a favourable (unfavourable) action by the party or by chance, finding that intention matters for unfavourable actions, but has only a weaker effect with favourable actions. Falk *et al.* (2000) compare responses to each of 13 hypothetical choices in a 'moonlighting game', where (with efficiency considerations) the first mover can either pass or take tokens from the responder, who can then respond by passing or taking. The data provide evidence of positive and negative reciprocity, comparing responses to intentional first-mover choices and random choices. In contrast, Cox and Deck (2005) find no evidence of either positive or negative reciprocity when comparing binary responses

---

[2] 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).

[3] To the best of our knowledge, this approach was first used in Brandts and Charness (2003).

[4] See also Falk *et al.* (2000) for a discussion of the role of intentions.

[5] For example, suppose a worker is assigned an intermediate wage of 50 in the game in Fehr *et al.* (1998). The worker can choose minimum (costless) effort and obtain (firm, worker) payoffs of (7, 50), or make a modest financial sacrifice and instead obtain payoffs of (14, 49) or (21, 48) or (28, 46) or (35, 44) etc.

[6] In an earlier study, Blount (1995) elicits 'minimum acceptable offers' in the ultimatum game, finding that the stated minimum depended on whether responders were told offers would be made by a self-interested party or generated by a random draw. However, as has been traditional in experiments by social psychologists, this study used deception, in that offers in fact were never forthcoming and all participants were actually paid a fixed fee.

in a mini-punishment game and a mini-'trust' game when a particular node is either chosen intentionally by a first mover or by a coin flip.[7,8]

A second approach involves comparing responses at a particular choice node in a game, according to how this choice node was reached. If behaviour at this node depends on the set of possible outcomes *off* the actual path of play, we move beyond a consequentalist framework to one where the context may matter, as foregone alternatives may yield indirect inferences regarding the intention of the first mover. The method used to test for intention is to systematically vary the foregone alternative(s). Brandts and Solà (2000) and Falk *et al.* (2003) use the 'mini-ultimatum game', where the proposer has only two options. The first study uses a between-subject design, where each individual responds to actual choices, while the second study uses a within-subject design, where each responder makes responses to hypothetical offers in four games where the available alternative offer is varied. They both find that people are more likely to reject a 20% offer in the ultimatum game when a 50% offer could have been chosen than when the foregone alternative was less than 20%, although results in the between-subjects design are generally weaker than results in the within-subjects design.[9]

Charness and Rabin (2002, 2005) compare behaviour at a particular node after the first mover intentionally foregoes a unilateral outside option (either 'favourable' or 'unfavourable' for the responder) and behaviour when the first mover had no choice. They find strong evidence of the relevance of intention, as negative reciprocity is observed; however, there is no evidence of positive reciprocity. Recognising that the responder has a number of plausible (and potentially self-serving) interpretations available concerning the first mover's intention, Charness and Rabin (2005) permit the first mover to express a preference between the responder's binary responses. Responder behaviour is sensitive to these expressed preferences and some evidence of positive reciprocity is found in this case.

While all of these previous studies employ a sensible approach for testing the importance of intention, there are nevertheless potential concerns with the methodology in each study. One such concern revolves around the possibility (mentioned above) that a responder is more generous when the other party had no choice than when this other party has made a deliberate (and favourable) choice. In many of the studies mentioned, one compares responses at a node reached intentionally to a node reached without the volition of the other party. This issue might obscure the effect of intention with regard to positive reciprocity, because we would not observe a higher rate of favourable responses to helpful intentional actions than to non-intentional choices unless positive reciprocity is strong enough to overcome this handicap. This issue might help to explain the often-contradictory evidence found in the several studies using this approach.

---

[7] On yet another hand, Cox (2000) finds evidence of positive reciprocity in a 'trust' game, as the amount returned depends on whether the amount received was sent intentionally or determined by an exogenous draw.

[8] In a slightly different vein, a study by Houser *et al.* (2005) explores sanctions and has a stochastic element. They find that whether a person is threatened intentionally by a self-interested party or randomly by nature does not matter, as the threat *per se* seems to be what arouses a response.

[9] Nelson (2002) also uses a within-subject, strategy-method design and finds that intentions matter.

Another feature present in many of the experimental studies cited above is the use of the *strategy method* (Selten, 1967): responders make contingent choices at every feasible decision node and the stated response to the actual first-mover choice is used for payoff determination. This provides a huge advantage in terms of data collection, particularly with respect to nodes that might only rarely be reached. And yet, it is possible that contingent responses vary systematically from actual responses,[10] in large part because the punishment or reciprocation experimenters are trying to study can be a visceral reaction that is not as likely to arise when people methodically consider hypothetical situations. The strategy method and the direct-response method sometimes yield different results (Güth *et al.*, 2001; Brandts and Charness, 2003) and sometimes yield similar results (Cason and Mui, 1997; Brandts and Charness, 2000). To the extent that tests consider *changes* in the rate of positive or negative responses rather than the level of the rate, the strategy method should be innocuous *if* it (plausibly) does not interact with the treatment status. In any case, if the level itself is important, one must be more cautious. Thus, the question of the effects of the strategy method is still not a settled issue.

In some of the studies on intention, a within-subject methodology is combined with the strategy method; here people make multiple contingent choices and only one is *ex post* payoff-relevant. On the one hand, within-subject designs may plausibly lead to spurious effects when respondents attempt to provide answers to satisfy their perceptions of the experimenter's expectations; this may be a particular concern when the responder must make a large number of contingent responses to the first mover's options. On the other hand, between-subjects designs, with each participant involved in only one treatment, inherently have substantial noise, and may miss important patterns. Thus, one chooses between concerns over obtaining potentially spurious effects and using less powerful tests.[11]

The motivation behind our design was to be able to make a comparison between responses made at a particular node that can be reached through two different deliberate choices by a first mover. Thus, we combine an intentional choice by a first mover with a stochastic process; we are unaware of any previous study using this approach. Our design permits comparisons across conditions with identical payoffs reached in different manners. We also wished to have intention be as clear as possible, and to ensure that workers knew the actual wage choice before making a decision. Thus, our approach can control for distributional issues and isolate the role of intention while staying clear of some potential pitfalls in the methodology used in previous experimental tests.

2. Experimental Design and Predictions

2.1. *Experimental Design*

Students at UC Berkeley and UC Santa Barbara participated in our experiments. Participants in Treatment 1 were students in an introductory organisational behaviour

---

[10] Roth (1995) was perhaps the first to make the point that forcing people to consider responses at every node might cause them to think differently.

[11] Some studies in economics have found substantial differences between with-subject and between-subject designs, as in the survey data in Gneezy (2005) and the experimental data in the working-paper version of Charness *et al.*, (2007). See Hsee *et al.* (1999) for a review of within- versus between-subject designs.

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.

Each session was conducted in a large classroom with two sides; half of the people were seated on each side. We flipped a coin in public 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.

The complete experimental instructions are provided in Appendix A. 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), with the amount chosen subtracted from the $12 endowment. The amount chosen was communicated as the chosen wage to each worker; we then flipped a coin separately in front of each worker to determine whether 'conditions' were good or bad, increasing or decreasing the wage by $2 when conditions were good or bad, respectively; the firm's payoff was unaffected by these conditions. All of this was common information.

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.[12] We show all of the resulting possibilities in Table 1.

There were two 'periods', where each firm received a new $12 endowment and was matched with a *different* worker the second time. As in the experimental instructions, we then chose one of these periods at random for actual payment. Each firm made its second choice without being informed of the worker's choice and the outcome from the first choice.

This design has the feature that a worker can receive a pre-response net payment of $6 from either a high wage or a low wage, so that we can test if the response is path-dependent. Nevertheless, if workers care (positively or negatively) about the firm's payoff in a distributional sense, we might expect the difference in the firm's

Table 1

Payoffs to the Firm and the Worker in Treatment 1

		Worker Effort Level*		
Wage paid by Firm	Conditions	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

* In end column the first number is the firm payoff and the second the worker payoff.

[12] 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 minimising strategic considerations (e.g., no repeated game considerations) and facilitating the analysis of the observed behaviour.

pre-response earnings in the two cases ($8 or $4, after a high or low wage) to affect effort even though the worker payoffs are the same.

To eliminate this consideration and to assess what difference intention makes to a worker when we hold constant both the worker and firm earnings, we conducted additional sessions with a slightly different design, in which conditions affected both parties. If 'wage conditions' were good, the firm paid an additional $2 to the worker; if these conditions were poor, the firm paid $2 less to the worker. To avoid the possibility of negative earnings, we increased the firm's endowment by $2 to $14. Thus, both a low wage and good wage conditions and a high wage and poor wage conditions led to pre-response payoffs of $8 for the firm and $6 for the worker. We show all of the resulting possibilities in Table 2.

We ran each group through two periods in order to gather more data, at the cost of non-independence of some observations. However, there is no scope for an individual reputation in our design and little scope for a 'group reputation' to form in just two periods (particularly since no feedback is given to the firm) so we would not expect strategic behaviour from workers.[13]

2.2. *Predictions*

In the past decade, laboratory observations of financial sacrifice have inspired behavioural-economic models of utility. There are two main classes of models: those that incorporate a role for intention and those that consider sacrifice to result from purely distributional motivations. In the category where intentions matter, Rabin (1993) and Dufwenberg and Kirchsteiger (2004) provide reciprocity models where one party considers the degree of kindness implicit in another party's choice and responds favourably to kindness but negatively to unkindness. Beliefs about intentions play a central role in the kindness-based models, with psychological game theory (Geanakoplos *et al.*, 1989) serving as the analytic tool. In the category of models without intentions, Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) presume that agents may sacrifice their own payoff to reduce disparities in relative payoffs; the Fehr

Table 2

Payoffs of the Firm and the Worker in Treatment 2

		Worker Effort Level*		
Wage paid by Firm	Wage Conditions	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

* In end column the first number is the firm payoff and the second the worker payoff.

[13] In some experiments with 10 or more multiple periods and anonymous matching (Brandts and Charness, 2003; Charness *et al.*, 2004; Charness and Kuhn, forthcoming), 'worker' behaviour unravels in the last couple of periods, suggesting that workers choose higher 'effort' in early periods in part to maintain a group reputation for effort provision and attract high wages.

and Schmidt model also presumes that people who draw the short stick may dislike inequality more than those who draw the long one.

More recently, models have combined intention-based reciprocity and a concern about distribution. Falk and Fischbacher's (2006) model provides a role for intention and uses Fehr and Schmidt's (1999) inequity aversion as the distributional underpinning. Charness and Rabin (2002) present a model that includes a distributional element comprised of a weighted average of the total payoffs and the minimum payoff; in addition, when one party 'misbehaves' with respect to the social standard, the weight assigned to that individual's payoff diminishes as others withdraw their concern for that party's welfare. One may also sacrifice money to reduce the material payoff of a misbehaving party, but the model does not include positive reciprocity.

The specific predictions for the Bolton and Ockenfels (2000), Fehr and Schmidt (1999), and Charness and Rabin (2002) models are detailed in Appendix B. To summarise, 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 a sufficient dislike of being behind (in their model, $\alpha > 1/3$ is the precise cutpoint for people to turn down \$1 if their counterpart would gain \$4; they report this is not an uncommon condition). There could be high or medium 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.

3. 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

Table 3

Treatment 1 *Results*

Wages	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 italicised rows pay identical wages.

© The Author(s). Journal compilation © Royal Economic Society 2007

Table 4

Treatment 2 Results

Wages	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 italicised rows pay identical wages.

well as the number of effort responses in each category for the wage offers and conditions. These data are also aggregated across high and low wages and poor and good conditions.

As can readily be seen, worker behaviour differs considerably across category in both Treatment 1 and Treatment 2. We summarise the patterns observed in our three main results:

RESULT 1. *The effort choice depends substantially on the chosen wage.*

In Treatment 1, workers punish the firm 22% of the time and reward the firm 9% of the time after a low wage is chosen; in contrast, workers punish 2% of the time, but reward 47% of the time after a high wage is chosen. In Treatment 2, workers punish the firm 41% of the time and reward the firm 7% of the time after a low wage is chosen; in contrast, workers punish 2% of the time, but reward 34% of the time after a high wage is chosen.

RESULT 2. *The path leading to a specific received wage affects the effort chosen.*

If only the worker's payoff matters, 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. However, we see that in both Treatment 1 and Treatment 2, worker behaviour is considerably different across these combinations (see the rows in italic). Punishment rates across these wage/condition combinations are substantially different in Treatment 1 (19% versus 3%) and Treatment 2 (40% versus 9%). Reward rates across these wage/condition combinations are quite different in Treatment 1 (3% versus 39%), although only slightly different in Treatment 2 (9% versus 14%).

RESULT 3. *Conditions do not significantly affect the effort chosen, except for the rate of reward after a high wage in Treatment 2.*

In both Treatment 1 and Treatment 2, punishment rates after a low wage are similar with poor conditions and good conditions, as are reward rates. After a high wage in both Treatment 1 and Treatment 2, there is no real difference in punishment rates across conditions after high wages, with a moderate difference in reward rates across conditions after high wages. However, there is a substantial difference in reward rates across conditions after high wages in both Treatment 1 (60% versus 39%) and Treatment 2 (58% versus 14%).

3.1. *Statistical Tests and Regressions*

We perform both nonparametric statistical tests and regressions to provide support for our comments above. Table 5 presents the comparisons mentioned in the three main results:

We see strong statistical support for our main results. When we correct for the fact that each worker makes two choices, by using individual (sometimes punish, never sacrifice, sometimes reward) tendencies, these test results are essentially unchanged.

Table 6 shows regressions on the determinants of effort, using effort choice as the dependent variable and wages and conditions as the explanatory variables. These regressions are generalised least squares regressions reflecting the three feasible effort choices (we use robust standard errors to account for the fact that observations are not independent for the two choices of a single worker).

The regressions in columns (1) and (2) confirm the patterns observed and described in Result 1 and Result 3, as a high wage induces substantially more effort than does a low wage, while conditions matter only in the interaction term in column (2). The

Table 5

Statistical Tests

	Comparison (n)	z-statistic, p-value*
Result 1		
Punishment rates, low wage vs. high wage, T1	15 of 67 vs. 1 of 51	$z = 3.21$, $p < 0.001$
Punishment rates, low wage vs. high wage, T2	35 of 85 vs. 3 of 41	$z = 3.88$, $p < 0.001$
Reward rates, high wage vs. low wage, T1	24 of 51 vs. 6 of 67	$z = 4.71$, $p < 0.001$
Reward rates, high wage vs. low wage, T2	14 of 41 vs. 2 of 85	$z = 5.02$, $p < 0.001$
Result 2		
Punishment rates, low/good vs. high/poor, T1	6 of 31 vs. 1 of 31	$z = 2.01$, $p = 0.044$
Punishment rates, low/good vs. high/poor, T2	16 of 40 vs. 2 of 22	$z = 2.57$, $p = 0.010$
Reward rates, low/good vs. high/poor, T1	3 of 31 vs. 12 of 31	$z = 2.67$, $p = 0.008$
Reward rates, low/good vs. high/poor, T2	1 of 40 vs. 3 of 22	$z = 1.71$, $p = 0.087$
Result 3		
Punishment, low wage, good vs. poor, T1	9 of 36 vs. 6 of 31	$z = 0.55$, $p = 0.582$
Punishment, low wage, good vs. poor, T2	16 of 40 vs. 2 of 22	$z = 0.21$, $p = 0.834$
Reward, high wage, good vs. poor, T1	12 of 20 vs. 12 of 31	$z = 1.49$, $p = 0.136$
Reward, high wage, good vs. poor, T2	11 of 19 vs. 3 of 22	$z = 2.98$, $p = 0.003$

*z-statistics from the difference of proportions (Glasnapp and Poggio 1985); tests are two-tailed.

Table 6

Determinants of Effort Random-effects Regressions, 3-choice Effort

	(1) Treatment 1	(2) Treatment 2	(3) Net wage = 6, Treatment 1	(4) Net wage = 6, Treatment 2
High wage	0.465***	0.386***	0.428***	0.407***
	(0.123)	(0.130)	(0.132)	(0.126)
Good conditions	0.067	−0.015		
	(0.117)	(0.109)		
High wage × Good conditions	0.168	0.544***		
	(0.193)	(0.189)		
Constant	0.860***	0.625***	0.903***	0.636***
	(0.090)	(0.083)	(0.098)	(0.084)
Observations	118	126	62	62
Number of Groups	59	63	46	47
Overall R^2	0.242	0.284	0.151	0.133

Standard errors in parentheses ***means significant at 1%

regressions in columns (3) and (4) confirm the claim in Result 2, as in both cases effort is substantially higher after a high wage and poor conditions compared to a low wage and good conditions.

3.2. *Discussion*

How do the predictions of the social-preference models fare? First, notice that the effect of the wage choice is quite strong and the effect of differences in distribution is typically small, suggesting that intention appears to be a stronger force than distribution in our data. Our featured comparison of high wage/poor conditions with low wage/good conditions provides clear support for the role of intentions. In fact, the results are consistent with the models of pure intentions-based reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004), except for the case of high wages and good conditions in Treatment 2, where the role of luck is strong. Consistent with the predictions of the Charness and Rabin (2002) model, the punishment rate is unaffected by conditions, while we see substantially more reward with good conditions than with poor conditions; however, not all of this model's parameters have been successfully estimated.

   As in Appendix B, the Bolton and Ockenfels (2000) distributional model predicts workers in both treatments will always choose medium effort after a low wage and poor conditions, and that medium effort will typically prevail in the other cases. However, this prediction is clearly rejected by our data, because punishment rates are as high as 42% after low wages and poor conditions, and reward rates are as high as 60%. The only prediction the Fehr and Schmidt (1999) model makes for all parameter values is for medium effort after high wages and poor conditions; in contrast, we see a 39% reward rate in Treatment 1 in these circumstances, although the parallel reward rate in Treatment 2 is only 14%. This model also predicts more punishment after a low wage with poor conditions than with good conditions, which is consistent with the slight trend in each treatment. Nevertheless, neither of these models can explain the significant difference in effort across the high/poor and low/good combinations in Treatment 2, where the interim firm and worker payoffs are identical.

A major difference between the two treatments is that in Treatment 1 the luck conditions affect only the workers' payoffs and do not appreciably affect effort choices, while in Treatment 2 a worker's bad luck is a firm's good luck. While the matter of luck does not seem to affect the urge to punish, there is greater reluctance to reward a firm for offering a high wage when the wage conditions are poor, where the firm has already been 'rewarded' by good luck at the worker's expense. This result is 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.

We feel that a firm's motivation for offering a high wage in our set-up is unlikely to be profit-maximisation, unlike the situation in many gift-exchange experiments; perhaps some firms prefer either equality or efficiency, considerations that could make a high wage optimal. Given the *ex post* outcomes, expected firm profits in Treatment 1 (Treatment 2) are $7.47 ($8.45) when the firm pays a low wage and $5.91 ($7.07) when the firm pays a high wage.[14] A firm concerned only with its own profits does better by choosing a high wage only when it expects that a low wage will be punished and a high wage rewarded; yet of the 56 workers who had the opportunity both to punish a low wage and to reward a high wage in the two periods, only six (just over 10%) did so. Thus, we believe that the intention behind a high wage is clear in our design.

4. 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.

An innovation 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 compare the effects of distributional payoffs and intentions readily, avoiding some of the methodological concerns present in past studies of the role of intention.

In fact, this feature of our design highlights an advantage of laboratory experiments over other forms of data: Only in a laboratory experiment can one arrange to have a test in which the outcome is precisely the same in two cases but the process leading to this outcome differs in these cases. We are certainly aware of the criticisms of laboratory experiments that have been made recently but we stress that there is considerable value in being able to cleanly identify qualitative effects (treatment effects); even some of the highly-critical papers (e.g., Levitt and List, 2005) acknowledge the strength of laboratory experiments in this regard. We favour a 'big tent' approach in which laboratory experiments, field experiments, and 'naturally-occurring' data are complementary, rather than mutually antagonistic; there is much to be learned from each method, and one hopes that these approaches all converge.

---

[14] This differs from many gift-exchange games, where the firms' multiplicative payoff function means that a firm can earn little without substantial worker effort, so that it turns out to be payoff-maximising for the firm to offer substantial wages. This confounds the interpretation of the motivation behind choosing a high wage offer.

We find that effort after a low wage does not depend much on distributional considerations; however, employee behaviour depends strongly upon the path that leads to the interim outcome. In contrast to some important recent models of social preferences (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), intentions appear to 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 behaviour in response to a high wage will be sensitive to external conditions. Some degree of rent-sharing exists where favourable conditions for the worker represent a loss for the firm.

Only models that combine distributional concerns with reciprocity motivations (Charness and Rabin, 2002; Cox *et al.*, 2005; Falk and Fischbacher, 2006) can possibly account for the observed patterns of effort choices. In this sense, our experimental results are consistent with those of previous studies such as Charness (2004), Charness and Rabin (2002), Cox (2000), Falk *et al.* (2003) and Offerman (2002).

Our findings on intentions' importance are consistent with the broader literature on procedural justice (Lind and Tyler, 1988).[15] That literature emphasises how most people like a system that will on average deliver good outcomes 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 emphasise 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.

4.1. *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 someone does you a favour but does it poorly. At the same time, assume that the favour creates an expectation that you will at some point do a costly favour in return. Here the initial favour, perhaps involving the best of intentions, makes you 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, as 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.

---

[15] Bruno Frey has substantial empirical work demonstrating the importance of procedural fairness, finding that the process does indeed matter. See for example Frey (1997).

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. 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 *et al.*, 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 predict behaviour 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. For example, company-wide sharing of gains may be important in providing high motivation for some workplaces – especially those relying on intrinsic motivation and group-based production – Levine (1995); Knez and Simester (2001). Future research is called for to understand how different settings influence the effects of absolute payoffs, relative payoffs, and intentions.

University of California Santa Barbara
University of California Berkeley

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 actual 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 actual 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 actual 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 actual 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 $\alpha \geq \beta$)
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 $\alpha \geq \beta$)
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:

$$U_i \equiv (1 - \lambda)\pi_i + \lambda\{\delta\min(\pi_i, \pi_j + bd_j) + (1 - \delta)[\pi_i + \max(1 - kd_j, 0)\pi_j] - fd_j\pi_j\}.$$

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

References

Adams, J. (1965). 'Inequity in social exchange', in (L. Berkowitz, ed.) *Advances in Experimental Social Psychology*, vol. 2, pp. 267–99, New York: Academic Press.

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

Blount, S. (1995). 'When social outcomes aren't fair: the effect of causal attributions on preferences'. *Organizational Behavior And Human Decision Processes*, vol. 63, pp. 131–44.

Bolton, G. and Ockenfels, A. (2000). 'ERC: A theory of equity, reciprocity, and competition'. *American Economic Review*, vol. 90, pp. 166–93.

Brandts, J. and Charness, G. (2000). 'Hot vs. cold: sequential responses in simple experimental games'. *Experimental Economics*, vol. 2, pp. 227–38.

Brandts, J. and Charness, G. (2003). 'Truth or consequences: an experiment', *Management Science*, vol. 49, pp. 116–30.

Brandts, J. and Solà, C. (2000). 'Reference points and negative reciprocity in simple sequential games', *Games and Economic Behavior*, vol. 2, pp. 227–38.

Brockner, J., Konovsky, M., Cooper-Schneider, R., Folger, R., Martin, C. and Bies, R. (1994). 'Interactive effects of procedural justice and outcome negativity on victims and survivors of job loss', *Academy of Management Journal*, vol. 37, pp. 397–409.

Caplan, R., Posner, K. and Cheney, F. (1991). 'Effect of outcome on physician judgements of appropriateness of care', *Journal of the American Medical Association*, vol. 265, pp. 1957–60.

Cason, T.N. and Mui, V. (1997). 'A laboratory study group polarization in the team dictator game', ECONOMIC JOURNAL, vol. 107(444) (September), pp. 1465–83.

Charness, G. (1996). 'Attribution and reciprocity in a simulated labor market', Chapter 2 of Ph.D. dissertation, University of California at Berkeley.

Charness, G. (2000). 'Responsibility and effort in an experimental labor market'. *Journal of Economic Behavior and Organization*, vol. 42, pp. 375–84.

Charness, G. (2004). 'Attribution and reciprocity in an experimental labor market', *Journal of Labor Economics*, vol. 22, pp. 626–47.

Charness, G. Fréchette, G. and Kagel, J. (2004). 'How robust is laboratory gift exchange?', *Experimental Economics*, vol. 7, pp. 189–205.

Charness, G., Haruvy, E. and Sonsino, D. (2007). 'Social distance and reciprocity: the internet vs. the laboratory', *Journal of Economic Behaviour and Organization*, vol. 63, pp. 88–103.

Charness, G. and Kuhn (forthcoming). 'Pay inequality, pay secrecy and effort: theory and evidence', *Journal of Labor Economics*.

Charness, G. and Levine, D. (2002). 'Changes in the fair employment contract: evidence from a quasi-experiment'. *Journal of Economic Behavior and Organization*, vol. 47, pp. 391–405.

Charness, G. and Rabin, M. (2002). 'Understanding social preferences with simple tests', *Quarterly Journal of Economics*, vol. 117, pp. 817–69.

Charness, G. and Rabin, M. (2005). 'Expressed preferences and behavior in experimental games', *Games and Economic Behavior*, vol. 53, pp. 151–69.

Cox, J. (2000). 'How to identify trust and reciprocity', *Games and Economic Behavior*, vol. 46, pp. 260–81.

Cox, J. and Deck, C. (2005). 'On the nature of reciprocal motives', *Economic Inquiry*, vol. 43, pp. 623–35.

Cox, J., Friedman, D. and Gjerstad, S. (2005). 'A tractable model of reciprocity and fairness', *Games and Economic Behavior*, vol. 59, pp. 17–45.

Dufwenberg, M. and Kirchsteiger, G. (2004). 'A theory of sequential reciprocity', *Games and Economic Behavior*, vol. 47, pp. 268–98.

Falk, A., Fehr, E. and Fischbacher, U. (2000). 'Testing theories of fairness – intentions matter', Working Paper No. 63, University of Zürich.

Falk, A., Fehr, E. and Fischbacher, U. (2003). 'On the nature of fair behavior', *Economic Inquiry*, vol. 41, pp. 20–6.

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

Fehr, E. and Gächter, S. (2000). 'Fairness and retaliation: the economics of reciprocity', *Journal of Economic Perspectives*, vol. 14, pp. 159–81.

Fehr, E., Kirchler, E., Weichbold, A. and Gächter, S. (1998). 'When social forces overpower competition: gift exchange in experimental labor markets', *Journal of Labor Economics*, vol. 16, pp. 324–51.

Fehr, E. and Schmidt, K. (1999). 'A theory of fairness, competition, and cooperation', *Quarterly Journal of Economics*, vol. 114, pp. 817–68.

Frey, Bruno. (1997). *Not Just For the Money. An Economic Theory of Personal Motivation*, Cheltenham: Edward Elgar.

Geanakoplos, J., Pearce, D. and Stacchetti, E. (1989). 'Psychological games', *Games and Economic Behavior*, vol. 1, pp. 60–79.

Glasnapp, D. and Poggio, J. (1985). *Essentials of Statistical Analysis for the Behavioral Sciences*, Columbus: Merrill.

Gneezy, U. (2005), 'Deception: the role of consequences', *American Economic Review*, vol. 95, pp. 384–94.

Güth, W., Huck, S. and Müller, W. (2001). 'The relevance of equal splits in ultimatum games'. *Games and Economic Behaviour*, vol. 37, pp. 161–9.

Houser, D., Xiao, E. McCabe, K. and Smith, V. (forthcoming), 'When punishment fails: research on sanctions, intentions and non-cooperation', *Games and Economic Behavior*.

Hsee, C., Blount, S., Loewenstein, G. and Bazerman, M. (1999). 'Preference reversals between joint and separate evaluations of options: a review and theoretical analysis', *Psychological Bulletin*, vol. 125, pp. 576–90.

Huang, P. (2000). 'Reasons within passions: emotions and intentions in property rights bargaining', *Oregon Law Review*, vol. 79, pp. 435–78.

Kahneman, D., Knetsch, J. and Thaler, R. (1986). 'Fairness as a constraint on profit-seeking: entitlements in the market', *American Economic Review*, vol. 76, pp. 728–41.

Knez, M. and Simester, D. (2001). 'Firm-wide incentives and mutual monitoring at Continental Airlines', *Journal of Labor Economics*, vol. 19, pp. 743–72.

Levine, D. (1995). 'Reinventing the workplace: How Business and Employees Can Both Win', Brookings Institution: Washington DC.

Levitt, S.D. and List, J.A. (2007), 'What do laboratory experiments measuring social preferences tell us about the real world', *Journal of Economic Perspectives*, forthcoming.

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

Morgenstern, A. (2004). 'Efficiency concerns and incentive provision – an experimental study', *Economics Letters*, vol. 83, pp. 335–41.

Nelson, W. (2002). 'Equity or intention: it is the thought that counts', *Journal of Economic Behavior and Organization*, vol. 48, pp. 423–30.

Offerman, T. (2002). 'Hurting hurts more than helping helps', *European Economic Review*, vol. 46, pp. 1423–37.

Rabin, M. (1993). 'Incorporating fairness into game theory and economics', *American Economic Review*, vol. 83, pp. 1281–302.

Roth, A. (1995). 'Bargaining experiments', in (J. Kagel and A. Roth, eds.) *Handbook of Experimental Economics*, Princeton, NJ: Princeton University Press.

Selten, R. (1967). 'Die strategiemethode zur erforschung des eingeschrankt rationalen verhaltens im rahmen eines oligopolexperiments', in (H. Sauermann, ed.), *Beitrage zur Experimentellen Wirtschafts-forschung*, pp. 136–168, Tubingen: J.C.B. Mohr.