Principal-Agent Settings with Random Shocks

Jared Rubin [a] and Roman Sheremeta [b]

[a] Argyros School of Business and Economics, Chapman University
One University Drive, Orange, CA 92866
[b] Department of Economics, Weatherhead School of Management,
Case Western Reserve University and the Economic Science Institute,
11119 Bellflower Road, Cleveland, OH 44106, U.S.A.

October 29, 2013

Abstract

Using a gift exchange experiment, we show that the ability of reciprocity to overcome incentive problems inherent in principal-agent settings is greatly reduced when the agent's effort is distorted by random shocks and transmitted imperfectly to the principal. Specifically, we find that gift exchange contracts without shocks encourage effort and wages well above standard predictions. However, the introduction of random shocks reduces wages and effort, regardless of whether the shocks can be observed by the principal. Moreover, the introduction of shocks significantly reduces the probability of fulfilling the contract by the agent, the payoff of the principal, and total welfare.

JEL Classifications: C72, C91, D63, D81, H50
Keywords: gift exchange, principal-agent model, contract theory, reciprocity, effort, shocks, laboratory experiment

---

Corresponding author: Roman Sheremeta, sheremet@chapman.edu

* We thank Ron Harstad, Brice Corgnet, Roberto Hernán-González, Gary Charness, Charlie Plott, David Rojo-Arjona, and seminar participants at Chapman University for helpful comments. We also thank Economic Science Institute and Chapman University for providing facilities and Koch Foundation for financial support. Any remaining errors are ours.

1. Introduction

This paper addresses two related sources of inefficiency that can arise in principal-agent relationships. First, a large literature notes that if the agent's effort is signaled imperfectly to the principal and monitoring is expensive or impossible, then it may be impossible to write a first-best contract, since the observed outcomes are not perfectly correlated with the agent's actions.[1] Second, if contracts are not exogenously enforceable or verifiable, endogenous enforcement through incentive compatibility requirements generally incentivizes agents to provide suboptimal levels of effort.[2] These two problems are related because it is impossible to exogenously enforce a contract (through legal or other institutions) which specifies effort requirements when effort is unobservable.

The unobservable effort problem is a common one for firms, as there are many types of tasks in which effort is positively correlated with observable outcomes, but these outcomes are also a function of random shocks (such as profits, number of sales, etc.). For example, the quantity of sales made by regional salespeople reflects both their effort and local demand fluctuations, where the latter are ostensibly random and difficult to observe. Hence, an employee can put in very little effort but perform well because of luck. Under these conditions, what is fair remuneration?[3] Should the employee be punished for lack of effort or rewarded for a good performance which predominantly came from luck? On the other hand, another employee can put forth very high effort but perform poorly because of bad luck. In that case, should the employee be punished for a bad outcome or rewarded for a high effort? Despite settled theoretical predictions, there is very little empirical research investigating how luck and effort play in remuneration in settings where effort is unobservable (Charness and Kuhn, 2011). This is understandable because it is difficult to measure empirically to what degree effort versus luck impacts individual performance. It is even more difficult to evaluate how employers reward

[1] See, for example, Harris and Raviv (1979), Holmström (1979), Shavell (1979), Holmström and Milgrom (1991), Baker (1992). Prendergast (1999) provides a more general overview of the contracts literature that emerged in the 1970s - 1990s.
[2] See, for example, Grossman and Hart (1983), Milgrom and Roberts (1992), and Laffont and Martimort (2002).
[3] According to the "informativeness principle" of Holmström (1979), when perfect information is not available, any observable measure of performance reveals information about the effort level chosen by the agent and should be used in the compensation contract. When effort is perfectly observable, the problem of optimal contract design is trivial: remuneration should be based on effort and not luck. This is sometimes referred to as the "accountability principle" (Konow, 2000, 2003), which states that remuneration should be based on the relevant variables that an individual can influence (i.e., effort) but not those that he cannot influence (i.e., luck).

effort versus luck, because remuneration is usually based on final performance which is a function of effort, ability and luck (Ericsson and Charness, 1994).

The second problem firms can face when contracting with employees is contract enforceability and verifiability. Even where legal institutions exist, writing a first-best, fully contingent contract is often impossible when effort is not verifiable. This problem is especially stark when random shocks affect the mapping from effort to outcome. For example, if a contract offers a wage in return for the first-best effort, the agent has incentive to provide less than the first-best effort if there is a high enough probability that he will get lucky (due to a positive production shock). Since the principal cannot verify whether the outcome is due to effort or luck, the principal cannot enforce the contract.

Fehr et al. (1997) provide experimental evidence that the contract enforceability problem is partially mitigated by behavioral concerns for reciprocity. They build on an extensive literature which suggests that the reciprocity motivation can help explain a host of results that are contrary to standard economic theory.[4] One implication of this literature is that contracts based on reciprocity come closer to the first-best than standard contract theory dictates. Fehr et al. (1997) test this implication with a gift exchange experiment, where principals offer contracts that include wages and desired effort levels. Agents who accept the contracts receive the wage and choose an effort level (where higher effort improves the principal's payoff), but they do not have to abide by the desired effort level in the contract. The money-maximizing Nash equilibrium is for the agent to provide zero effort (since it is costly and they cannot be punished) and for the principal to thus offer the lowest possible wage. In their experiment, however, agents frequently show positive reciprocity; not only do they provide more effort than the money-maximizing Nash equilibrium prediction, but their effort level is increasing in the wage offered by the principal. These results are exacerbated when principals are also allowed to exhibit reciprocity. In one treatment, Fehr et al. (1997) introduce a third stage in which principals can pay to punish or reward the agent after observing their effort. Although the addition of this stage does not alter the money-maximizing Nash equilibrium predictions of wage or effort, they find that allowing

[4] There is a wealth of experimental evidence that both positive reciprocity and negative reciprocity have important effects on actions, with negative reciprocity being shown as more salient. In the context of the gift-exchange experiment employed in this paper, see Charness and Haruvy (2002), Charness (2004), Charness and Dufwenberg (2006), Fehr and Schmidt (2007), and Houser at al. (2008). Rabin (1993) provides the canonical model introducing reciprocity into game theory, and Falk and Fischbacher (2006) provide a theory connecting the reciprocity motive to a host of standard experimental results. Fehr and Gächter (2000) provide a survey of the literature on reciprocity.

both sides to exhibit reciprocity significantly increases effort (and thus efficiency), and that both principals and agents are better off than they are when only agents are allowed to show reciprocity. Fehr et al. (2007) provide further evidence that this type of bonus contract vastly outperforms standard incentive-based contracts despite relying on unenforceable actions.

These papers contribute significantly to our knowledge of how behavioral incentives encourage contract enforcement in the absence of explicit incentives. Yet, Fehr et al. (1997) and Fehr et al. (2007) only consider how reciprocity improves contract efficiency under perfect information. In their experiments, principals can reward or punish agents based on perfectly observed effort – there are no random shocks affecting the mapping from effort to outcome. This is an important omission, because the types of contracts they are concerned with are often difficult to enforce in the real world *precisely because* outcomes are affected by shocks and thus optimal effort levels are impossible to induce in an exogenously enforced contract. Indeed, it is not clear ex-ante how the introduction of shocks interacts with the reciprocity motive. Do principals exhibit reciprocity when they are unsure that the outcome which they observe is the result of the agent's effort?

This paper addresses this problem. We conduct a gift exchange experiment similar to Fehr et al. (1997), except that the principal receives an imperfect signal of the agent's effort in some treatments. Our first treatment is similar to Fehr et al.'s (1997) bonus treatment. Principals and agents are randomly matched and the principal offers a wage and asks for a desired effort. The agent then receives the wage and can choose any effort (where the cost of effort is increasing in effort chosen). The principal can then reward or punish the agent, although either is costly. There are no shocks in this treatment, so we employ it as our baseline. The second treatment is exactly the same as the first, except that we add a random (uniformly distributed) number to the agent's effort. In this treatment, there is still perfect information; the principal observes both the effort level *and* the random number when making her decision of how much to punish or reward the agent. The final treatment is exactly like the second treatment, except that principals only observe the outcome (effort + random number), not the agent's effort. Relationships in all treatments are one-shot and anonymous, so reputational concerns are absent.

Consistent with previous literature on gift exchange (Fehr et al., 1997, 2007; Charness and Kuhn, 2011), we find that bonus contracts without shocks encourage effort and wages well above standard predictions. However, we also find evidence that this result is partially mitigated when random shocks are present. The mere introduction of shocks reduces wages and effort, *regardless of whether the shocks are observed by the principal*. If it is solely reciprocity encouraging high wages and effort, there should be no difference between the baseline treatment (without shocks) and the treatment where shocks are perfectly observed, since the reciprocity motive is based on the other's action, not the outcome emanating from the action.[5] Yet, this is not what we find. Wages and effort are significantly lower in treatments where shocks are perfectly observed relative to the baseline. Moreover, we observe no differences in behavior between treatments where shocks are present and observable and treatments where shocks are present and unobservable.

These results have important implications regarding the ability of behavioral responses to help mitigate principal-agent problems inherent in contract design. These results by no means deny that reciprocity can support greater cooperation than equilibrium predictions; indeed, even with the introduction of random shocks, effort and efficiency are well above the money-maximizing Nash equilibrium prediction. Our results do indicate, however, that the effects of reciprocity are partially mitigated by the information available to subjects and the manner in which intentions are related to observable outputs. In particular, these results suggest that the reciprocity motive is based on the outcome (i.e., which is a function of effort and shocks) of others' actions and not simply on their intentions (i.e., effort).

2. Experimental Design and Procedures

Our experimental design is built on a variation of a gift exchange game. The game consists of three stages. In stage 1, the principal offers contract (w,$\underline{e}$) to the agent. That is, the principal offers a wage w (any integer number between 1 and 100) and the desired effort $\underline{e}$ (an integer number between 0 and 14) that she would like the agent to undertake.[6] In stage 2, the agent receives the wage w and chooses an effort level e, which does not have to be equal to the desired effort $\underline{e}$ specified by the contract. The cost of effort $c(e)$ is an increasing and convex

[5] There are several recent experiments documenting that subjects often make their decisions based on both outcomes and intentions (actions) of others (Falk and Fischbacher 2006; Charness and Levine, 2007; Erkal et al., 2011; Rey-Biel et al., 2012; Gurdal et al., 2012; Cappelen et al., 2013).

[6] We chose the range between 0 and 14 for effort and desired effort for two reasons. First, it ensures that the maximum cost of effort is less than the maximum possible wage (cost of effort of 14 is equal to 98). Second, we wanted the efficient effort (10) to be an internal point (between 0 and 14), so that agents were not anchored to the efficient effort artificially (which could happen if the effort range was between 0 and 10).

function of effort, where $c(e) = e^2/2$. In stage 3, the principal first observes the outcome $y = e + \varepsilon$, which is a function of effort e and a uniformly distributed random component ε (an integer number between -2 and +2). As we will explain below, the primary difference between treatments is what the principal can observe ($\{y, e, \varepsilon\}$ or just y). After observing y, the principal chooses an adjustment level a (an integer number between -5 and +5), which can be either in a form of a bonus ($a > 0$) or punishment ($a < 0$).[7] The payoff of the principal is $\pi^P = 10y - w - |a|$ and the payoff of the agent is $\pi^A = w - c(e) + 10a$.

We employ three treatments, which we name based on what the principal observes. In the baseline Effort-Only treatment there is no random component (i.e., $\varepsilon = 0$), and the principal directly observes effort e (there is no difference between effort and outcome, since $y = e$). This treatment is similar to Fehr et al.'s (1997) "bonus" treatment and provides a baseline to which we compare our results. In the Effort-Shock treatment, there is a random shock component ε, which the principal observes. That is, the principal directly observes y, e, and ε. Finally, the Outcome-Only treatment is the same as the Effort-Shock treatment, but the principal only observes outcome y and does not know the composition of y.[8]

In all treatments, the money-maximizing subgame perfect equilibrium is for the agent to choose an effort of zero (i.e., $e = 0$) and for the principal to make an adjustment of zero (i.e., $a = 0$). The socially optimal actions are for the agent to choose an effort of 10 (i.e., $e = 10$) and for the principal to provide an adjustment level of +5 (i.e., $a = 5$).

We recruited subjects randomly from the student body of a mid-sized university in the United States. A total of 216 subjects were recruited from a standard campus-wide subject pool. Subjects interacted with each other anonymously over a local computer network. The experiment, which lasted an average of 45 minutes total, proceeded as follows. Upon arrival, subjects were randomly assigned to computer terminals and received instructions (see Appendix) corresponding to one of the three treatments. The experiment was computerized using z-Tree

[7] We chose the range between -5 and 5 for the adjustment because we wanted the ability to punish or reward to be large enough that most subjects would choose an internal point (to reduce censoring biases). We felt that this range accomplishes both of these goals while not being so large that contracts are completely based on bonuses or punishments.

[8] The two treatments that we introduce are novel and have not been studied previously. However, some elements of our design are related to Xiao and Kunreuther (2012), who examine behavior in a two-person prisoner's dilemma game with stochastic versus certain outcomes, and to Cappelen et al. (2013), who study fairness views about risk-taking with ex ante versus ex post stochastic outcomes.

(Fischbacher, 2007). We ran 9 sessions (3 sessions per treatment) with 24 subjects in each session.

Within each session, subjects were split into 3 groups of 8.[9] Within each group of 8, 4 subjects were assigned to be principals and 4 were assigned to be agents. Subjects stayed in their role assignment throughout the entire experiment. In each session there were 10 periods of play. In each period subjects from opposite role assignments were randomly matched to form a principal-agent pair. After each period subjects were randomly re-matched with someone of the opposite role assignment within their 8-person group to form a new principal-agent pair. Each period proceeded in three stages. In the first stage, the principal chose a reward (an integer number between 0 and 100) and a desired effort (an integer number between 0 and 14) for the agent. After observing the reward and the desired effort, in the second stage, the agent chose an effort level (an integer number between 0 and 14). To determine the outcome, in the Outcome-Only and Effort-Shock treatments, the computer added to the effort a randomly selected number (an integer between -2 and +2).[10] Then, depending on the treatment, the computer displayed to the principal either only the outcome (Outcome-Only), the outcome, effort, and the random number (Effort-Shock), or effort (Effort-Only). After observing the relevant information, in the third stage, the principal choose an adjustment level for the agent (an integer between -5 and +5).

At the end of each experiment, 1 out of 10 periods were randomly selected for payment.[11] The earnings in this period were exchanged at rate of 10 francs = $1. All subjects also received a participation fee of $20 to cover potential losses. On average, subjects earned $26 each (maximum $42 and minimum $7), which was paid anonymously and in cash.[12]

[9] We divided the subjects into three groups per session in order to have three independent observations at the session level. This allows for the use of non-parametric tests, which we employ in Section 3. Subjects were not told that they were split into three groups of eight.

[10] We allowed principals to receive a negative payout if effort plus the random number was negative.

[11] When subjects are paid for multiple periods in a single experiment, the payment from one period may impact subjects' choices in another. According to Azrieli et al. (2013), paying for one random selected period is the only incentive compatible payment mechanism.

[12] The fact that subjects receive a high participation fee of $20 does not diminish the saliency of subject payments, because subjects may win or lose a substantial amount of money. In fact, in our experiment, some subjects made as much as $42, while others made as little as $7.

3. Results

We focus on contracts that satisfy individually rational and incentive compatibility requirements (IR/IC). These are contracts in which both the principal's and the agent's payoffs are non-negative, conditional on the contract being fulfilled. Specifically, we focus on all (w,$\underline{e}$) contracts, such that $10\underline{e} - w \geq 0$ and $w - c(\underline{e}) \geq 0$.[13] Out of 2160 contracts observed in our experiment, 1338 (62%) can be classified as satisfying IR/IC.[14] Table 1 provides the summary statistics across all three treatments, using only IR/IC contracts (top panel) and all contracts (bottom panel). In Sections 3.1-3.5, we focus on contracts satisfying IR/IC. Yet, all major results hold when we combine IR/IC and non-IR/IC contracts together. We analyze in detail non-IR/IC contracts in Appendix B.

When performing statistical tests, we mainly use non-parametric tests to examine treatment effects. Each treatment has a total of 9 independent observations (72 subjects per treatment, split into 9 separate groups of 8 subjects each). When appropriate, we also estimate panel models with individual subjects representing random effects (to control for individual effects), standard errors clustered at the single re-matching group level of 8 subjects (to control for possible correlation within a matching group), and an inverse period trend (to control for learning and experience). We consider the results starting with stage 3 first and work our way backwards to stage 1.

3.1. Adjustment

In stage 3, principals choose an adjustment after seeing either the effort of the agent (in Effort-Only and Effort-Shock) or the outcome (in Outcome-Only and Effort-Shock) in stage 2. Figure 1 displays the average adjustment by treatment, while Figure 2 displays the distribution of adjustment by treatment. Both the distribution and the average adjustment levels are very similar in all three treatments. Based on the Wilcoxon rank-sum test there is no significant difference in the adjustment level between treatments: Effort-Only versus Outcome-Only (0.26 versus -0.62; p-value = 0.11, $n_1 = 9$, $n_2 = 9$), Effort-Only versus Effort-Shock (0.26 versus -0.51; p-value =

[13] We chose not to put any restrictions on the principal's decisions, because some ex-ante "non-IR/IC" contracts (w,$\underline{e}$) may be IR/IC ex-post, given a certain level of an adjustment a. For an experiment where the principal can only offer contracts which satisfy IR/IC see Bartling et al. (2012).

[14] The number of contracts which satisfy IR/IC is very similar across the three treatments (i.e., 64% in the Effort-Only, 59% in the Outcome-Only and 62% in the Effort-Shock treatment).

0.20, $n_1 = 9$, $n_2 = 9$), and Outcome-Only versus Effort-Shock (-0.62 versus -0.51; p-value = 0.96, $n_1 = 9$, $n_2 = 9$).[15]

This suggests the possibility that the adjustment mechanism works relatively similar in all three treatments. However, these results may arise from the fact that we consider the *unconditional* adjustment in stage 3; if the first two stages are different in the two treatments and principals condition their adjustment on actions taken in the previous stages, then focusing on the unconditional adjustment obscures the determinants of the adjustment across treatments. Indeed, if the reciprocity motive is present in the principal's decision, we expect the adjustment to be a function of how "kindly" she was treated by the agent in stage 2. In other words, we expect the principal's adjustment to be a function of the difference between the effort (or outcome) she observes in stage 2 minus the desired effort proposed in stage 1. It is also possible that the principal expects the agent to show reciprocity in stage 2 if the principal gives a large wage in stage 1, so the adjustment may also be conditional on wage.

We test whether principals condition their adjustments based on previous actions. Table 2 reports the estimation results of different panel models where individual subjects represent the random effects, and the standard errors are clustered at the single re-matching group level. The dependent variable in all specifications is the *adjustment* and the independent variables in specifications (1)-(3) are an inverse of a *period* trend, *wage*, *effort – desired effort* (in Effort-Only and Effort-Shock), and *outcome – desired effort* (in Outcome-Only).[16] In specifications (1) and (2), *adjustment* is positively correlated with *effort – desired effort*. In specification (3), *adjustment* is positively correlated with *outcome – desired effort*. This finding supports the idea that principals show reciprocity, since they reward higher effort (outcome) relative to desired effort. Also note that there is a positive correlation between *adjustment* and *wage* (although not significant in specification 3). One interpretation is that principals use *adjustment* and *wage* as complements.

While these results suggest that principles reward a "kind" effort with kindness of their own, the magnitude of this reward is different across treatments. In the Effort-Only treatment, principals increase their average adjustment by 0.78 for every unit of effort given (relative to

[15] Based on the Wilcoxon matched-pairs signed-rank test, we also find that the absolute value of adjustment in all three treatments is significantly higher than the money-maximizing Nash equilibrium prediction of $a = 0$ (all p-values < 0.01).

[16] Principals do not see effort in the Outcome-Only treatment, so we condition on outcome minus desired effort.

desired effort), whereas the marginal increase is only 0.45 in response to an increase in effort in the Effort-Shock treatment. In these two treatments, principals see the *same information*. The difference in the magnitude of these coefficients therefore suggests that principals do not reciprocate based solely on the intention (i.e., effort) of the agent. Indeed, in specification (4), we also include the *shock* as an independent variable. We find that the *adjustment* and the *shock* variables are positively correlated, suggesting that principals punish or reward agents based on both *intentions* and *outcomes*. In fact, a comparison of specifications (3) and (5) suggests that principals respond to an increase in *outcome* the same (0.44) regardless of whether or not the effort is observed.[17]

Result 1: There is no significant difference in the unconditional *adjustment* level between treatments. The adjustment level varies positively with effort and outcome. The magnitude of the change in adjustment in response to an increase in effort is weaker when shocks are present. The magnitude of the change in adjustment in response to an increase in outcome is the same regardless of whether effort is observed or not.

3.2. Effort

We next consider the effort that the agent chooses in stage 2. Figure 3 displays the average effort by treatment, while Figure 4 displays the distribution of effort by treatment. Based on the Wilcoxon rank-sum test, we find that the average unconditional effort in the Effort-Only treatment is higher than in the Outcome-Only treatment (6.86 versus 5.08; p-value < 0.01, $n_1 = 9$, $n_2 = 9$) and the Effort-Shock treatment (6.86 versus 5.07; p-value < 0.01, $n_1 = 9$, $n_2 = 9$). On the

[17] It is possible that the relationship between "reciprocity" (adjustment) and "kindness" (effort gap or outcome gap) is not linear (Baumeister et al., 2001; Offerman, 2002; Andreoni et al., 2003; Charness, 2004; Bellemare et al., 2007). Bellemare et al. (2007), for example, suggest that reciprocity is a concave function of kindness (i.e. increasing in the degree of kindness increases reciprocity, but at a diminishing rate). Moreover, following the seminal paper by Baumeister et al. (2001), many studies have shown that "negative" reciprocity is stronger than "positive" reciprocity (also see Charness and Kuhn 2011). In results that are available upon request, we control for both non-linearities and distinctions between positive and negative kindness. Our results indicate that both positive and negative reciprocity increase in the degree of "kindness" (i.e., *effort – desired effort* and *outcome – desired effort* are positively correlated with *adjustment*). Moreover, positive reciprocity increases at a diminishing rate (i.e., $(\textit{effort – desired effort})^2$ and $(\textit{outcome – desired effort})^2$ are negative) and negative reciprocity decreases at a diminishing rate (i.e., $(\textit{effort – desired effort})^2$ and $(\textit{outcome – desired effort})^2$ are positive). These results are not always statistically significant, however.

other hand, the average effort is not different between the Outcome-Only and Effort-Shock treatments (5.08 versus 5.07; p-value = 0.99, $n_1 = 9$, $n_2 = 9$).

There are two reasons why agents may choose effort greater than the money-maximizing Nash prediction of zero. First, they may believe that a higher effort will lead to a greater reward (or smaller punishment) in stage 3. We showed in the previous section that such beliefs are accurate, although there are treatment differences. Second, they may exhibit positive reciprocity if the principal gives them a high wage in the first stage. That is, their effort is in part *conditional* on actions taken in stage 1. We test this possibility by conducting a panel analysis within each treatment. Table 3 reports the estimation results of different panel models, where the dependent variable in all specifications is the subject's *effort* and the independent variables are an inverse of a *period* trend, *wage* and *desired effort*. In all specifications, there is a positive and significant relationship between *wage* and *effort* (although it is only significant at the 10% level in the Effort-Shock and Outcome-Only specifications), suggesting a gift exchange story between the principal and the agent.

Result 2: There is a greater *effort* in the Effort-Only treatment than in the other two treatments, while there is no significant difference in effort between the Effort-Shock and Outcome-Only treatments. The effort level responds positively to wage in all three treatments.

It is reasonable to suspect that the principal's willingness to reciprocate is not just a function of the absolute level of *effort* (or *outcome*), but it is also a function of the difference between *effort*/*outcome* and *desired effort*. Indeed, the results in Table 3 indicate that the magnitude of the effect of *desired effort* on the *effort* chosen differs between treatments. In the Effort-Only treatment, principals receive 61% of each additional unit of *effort* they desire, whereas the magnitude is 41% and 12% in the Effort-Shock and Outcome-Only treatments, respectively. These results suggest that agents form reasonably correct beliefs regarding how principals will act in the adjustment period. We showed in Result 1 that the adjustment response to *effort – desired effort* is strongest in the Effort-Only treatment, indicating that agents with correct beliefs should increase their effort the most in this treatment in response to an increase in desired effort.

Yet, even if agents correctly predict how principals act in stage 3, it is not clear ex ante how the introduction of shocks affects effort relative to desired effort. First, if agents believe that the most important thing to principals is *whether* the contract was fulfilled (i.e., effort/outcome ≥ desired effort) rather than by how much it was fulfilled by, we should expect to see the vast majority of effort within the interval [-2, 2] of desired effort. Any effort lower than this range allows the principal to know with 100% probability that the agent did not fulfill the contract, while any effort higher than this range involves more costly effort without affecting the principal's perceived probability that the contract was fulfilled. This is precisely what we find in Figure 5, which shows the distribution of "effort – desired effort" in all three treatments. This figure indicates that the vast majority of observations in all three treatments fall within the interval [-2, +2], suggesting that agents do not perceive the desired effort simply as a cheap talk but rather as a concrete indication of the principal's expectations.

It is also quite clear from Figure 5 that the distribution of effort – desired effort is different in the three treatments. What can explain this? If agents are very risk averse, they may choose to give more effort than desired effort in the Outcome-Only treatment in order to avoid any chance of being perceived as underperforming the desired effort. On the other hand, this logic indicates that agents are more likely to choose effort greater than the desired effort in the Outcome-Only treatment relative to the Effort-Only treatment. Whether agents choose effort greater than desired effort more frequently in the Outcome-Only treatment relative to the Effort-Shock treatment depends on the degree to which agents believe that principals reward/punish based on *outcome* relative to *intention*. We do indeed find that the probability of effort exceeding the desired effort (based on the Wilcoxon rank-sum test) is significantly lower in the Effort-Only treatment than in the Outcome-Only treatment (0.07 versus 0.19; p-value = 0.03, $n_1 = 9$, $n_2 = 9$), but there is no significant difference between the Effort-Only and Effort-Shock treatments (0.07 versus 0.14; p-value = 0.28, $n_1 = 9$, $n_2 = 9$) or between the Outcome-Only and Effort-Shock treatments (0.19 versus 0.14; p-value = 0.59, $n_1 = 9$, $n_2 = 9$).

Meanwhile, if agents at all suspect that principals base their adjustments on outcome rather than effort, we should expect to see contracts be *exactly* fulfilled (i.e., effort is equal to desired effort) in the Effort-Only treatment relative to the other two treatments. There are no shocks in this treatment, so effort is equal to outcome. Our results confirm this intuition. Agents choose efforts exactly specified by the contract in the Effort-Only treatment significantly more often than in the Outcome-Only treatment (0.41 versus 0.12; p-value < 0.01, $n_1 = 9$, $n_2 = 9$) and the Effort-Shock treatment (0.41 versus 0.23; p-value = 0.03, $n_1 = 9$, $n_2 = 9$). There is no difference between the Outcome-Only and Effort-Shock treatments (0.12 versus 0.23; p-value = 0.10, $n_1 = 9$, $n_2 = 9$), although the difference nears statistical significance.

Result 3: Effort levels respond positively to *desired effort* in the Effort-Only and Effort-Shock treatments. Yet, there is a greater probability that effort *exceeds desired effort* in the Outcome-Only treatment than in the Effort-Only treatment, while there is a greater probability that the contract is *exactly fulfilled* in the Effort-Only treatment than in the other two treatments.

3.3. Wage and Desired Effort

In terms of welfare, the most important result presented thus far is Result 2, which indicates that effort is greater in the Effort-Only treatment than in the other two treatments. Where does this extra effort come from? Result 3 indicates that it does not come from agents giving extra effort relative to desired effort, although it may come from agents giving less effort than desired less frequently. This leaves two, non-mutually exclusive, possibilities: (i) principals give higher wages and/or, (ii) ask for higher desired efforts in stage 1.

Figure 6 displays the average wage and desired effort by treatment, while Figure 7 displays the distribution of wage and desired effort by treatment. The average wage and desired effort are the highest in the Effort-Only treatment. Using the average within a single re-matching group over all periods as one independent observation, the Wilcoxon rank-sum test shows that the average wage in the Effort-Only treatment is significantly higher than in the Effort-Shock treatment (46.27 versus 38.87; p-value = 0.05, $n_1 = 9$, $n_2 = 9$) and Outcome-Only treatment (46.27 versus 37.36; p-value = 0.02, $n_1 = 9$, $n_2 = 9$). Similarly, we find that the average desired effort is higher in the Effort-Only treatment than in the Effort-Shock treatment (8.13 versus 6.91; p-value < 0.01, $n_1 = 9$, $n_2 = 9$) and the Outcome-Only treatment (8.13 versus 6.65; p-value < 0.01, $n_1 = 9$, $n_2 = 9$). On the other hand, wage and desired effort are not different between the Outcome-Only and Effort-Shock treatments (p-values = 0.89 and 0.63, respectively).

Result 4: There is a greater *wage* and *desired effort* in the Effort-Only treatment than in the other two treatments, while there is no significant difference in wage and desired effort between the Effort-Shock and Outcome-Only treatments.

Result 4 indicates that the higher effort level observed in the Effort-Only treatment in Result 2 results from both higher wages and higher desired effort levels in the Effort-Only treatment. Why do principals offer a higher wage and ask for greater desired effort in the Effort-Only treatment? Part of the answer follows from Table 3, which indicated that effort responds more strongly to desired effort in the Effort-Only treatment than in the other two treatments (0.61 versus 0.41 and 0.12, comparing the coefficients on *desired effort* in all three specifications). Hence, principals have more to gain from asking for higher desired effort in the Effort-Only treatment than in the other two treatments. If higher wages are necessary to induce such effort, this would also explain why wage is greater in the Effort-Only treatment. In fact, we find that there is a strong correlation between wage and desired effort, $\rho = 0.85$, indicating that higher wages are associated with higher desired effort.

3.4. Payoffs and Welfare

As a consequence of higher wage and higher effort, the Effort-Only treatment generates significantly higher payoff to the principal than the Outcome-Only treatment (20.39 versus 9.17; p-value = 0.03, $n_1 = 9$, $n_2 = 9$) and the Effort-Shock treatment (20.39 versus 12.09; p-value = 0.05, $n_1 = 9$, $n_2 = 9$). Yet, the principal's payoff is not significantly different between the Outcome-Only and Effort-Shock treatments (9.17 versus 12.09; p-value = 0.57, $n_1 = 9$, $n_2 = 9$). When comparing payoffs of agents, we find no significant differences between the three treatments (all p-values > 0.45).

Result 5: *Principals' payoffs* in the Effort-Only treatment are higher than in the other two treatments, while there is no significant difference between the Effort-Shock and Outcome-Only treatments. There is no statistically significant difference in the *agents' payoffs* between any of the treatments.

The fact that principals are better off in the Effort-Only treatment but agents are not suggests that although principals offer higher wages in the Effort-Only treatment, this translates into higher effort levels which leave the agents equally well off but make principals better off. The principals are made better off by enough in the Effort-Only treatment that the *overall* welfare (principal's payoff + agent's payoff) is greater in the Effort-Only treatment than in the other two treatments: Effort-Only versus Outcome-Only (39.30 versus 31.58; p-value = 0.02, $n_1 = 9$, $n_2 = 9$) and Effort-Only versus Effort-Shock (39.30 versus 29.44; p-value < 0.01, $n_1 = 9$, $n_2 = 9$). On the other hand, there is no significant difference in the total welfare between the Effort-Shock and Outcome-Only treatments (29.44 versus 31.58; p-value = 0.57, $n_1 = 9$, $n_2 = 9$).

Result 6: *Total welfare* is greater in the Effort-Only treatment than in the other two treatments, while there is no significant difference between the Effort-Shock and Outcome-Only treatments.

4. Discussion and Conclusion

We conduct a gift exchange experiment in which the agent's outcome depends on both effort and luck. Consistent with the previous literature on gift exchange (Fehr et al., 1997, 2007; Charness and Kuhn, 2011), we find that bonus contracts without a shock component encourage effort and wages well above the money-maximizing Nash equilibrium prediction. We also find that a significant number of agents do not shirk and exert at least as much effort as is specified by the contract.

Two fundamental findings follow from our results. The first finding is that people reward on the basis of both *intentions* and *outcomes*. This is not a new result. For instance, Falk and Fischbacher (2006) provide a theory of reciprocity centered on the idea that people base reciprocity on both the intentions and consequences of an action. Likewise, our result is consistent with a large literature on retrospective voting that finds voters reward/punish politicians based on outcomes over which politicians have no control (Healy et al., 2010; Gasper and Reeves, 2011). It is also consistent with a large literature in psychology on outcome bias (Baron and Hershey, 1988; Marshall and Mowen, 1993; Mazzocco et al., 2004). There are also recent studies on risk taking, redistribution and charitable giving that show that some subjects condition their giving and reciprocity on both effort and luck of others (Charness and Levine, 2007; Erkal et al., 2011; Rey-Biel et al., 2012; Gurdal et al., 2012; Cappelen et al., 2013).

The novel and important result of our study is that the introduction of a shock in the principal-agent settings significantly reduces wages and effort, *regardless* of whether the shock can be observed by the principal.[18] The introduction of shocks in the principal-agent settings also significantly reduces the probability of fulfilling the contract by the agent, the payoff of the principal, as well as the total welfare. The fact that shocks, even perfectly observable, have such a significant and perhaps unexpected effect in principal-agent settings has important implications for the design of optimal contracts.

What is it about the addition of shocks – observed or unobserved – that encourages principals to offer contracts with lower wages and desired effort levels? Why does the addition of shocks make agents less responsive to desired effort? While we cannot pinpoint the exact behavioral mechanism underlying our results, we *can* say something about theories which are inconsistent with our results. In particular, a satisfactory theory must account for the fact that the *observability* of the shock does not affect effort or welfare. This suggests that are results are not being driven by agents "hiding behind randomness", where they give less effort when confronted with a good shock. For instance, Andreoni and Bernheim (2009) argue that people like to be perceived as fair, and thus act selfishly when they can ascribe their actions to chance.[19] If this motivation were driving our results, there would be less effort given in the Outcome-Only treatment (where principles do not see the shock) than in the Effort-Shock treatment (where principles do see the shock). Indeed, any explanation based on the observability of actions cannot explain our treatment differences.

What, then, can explain our results? We believe that there are two, non-mutually exclusive conjectures that are consistent with our results. The first conjecture has to do with the nature of gift exchange. Specifically, as wage and desired effort levels increase, the downside risk becomes greater due to the gift exchange nature of the game – the agent may not choose the

[18] Our findings contrast with the findings of Sloof and Van Praag (2010), who document that subjects exert higher efforts when there is more noise in the production process. However, our results are not directly comparable since we examine behavior of subjects in a chosen-effort principal-agent setting, while Sloof and Van Praag (2010) examine behavior in a real-effort experiment without a principal.

[19] Aimone and Hauser (2011) also show that the "betrayal aversion" impulse is weaker when agents can hide behind randomness. In their experiment, betrayal aversion induces greater trust and hence greater efficiency. As noted above, however, this cannot explain why we do not find differences between our Outcome-Only and Effort-Shock treatments.

desired effort level, and thus the higher wage is wasted. Likewise, when agents choose higher efforts, the downside risk that the principal will not reciprocate in the third stage is greater, since the effort chosen is more costly both in absolute and marginal terms. As the costs increase, players must be compensated by either higher payouts or lower risk. The Effort-Only treatment offers the lowest uncertainty of the three, since agents know that principals receive an amount corresponding exactly to the amount of effort that they give. In this treatment, agents do not have to be concerned about whether the principal rewards based on intention or outcome. This, in turn, allows higher levels of effort to be sustained, as the additional risk inherent in the other two treatments makes high levels of effort too costly to be worth the risk.

Second, it may be the case that the factors affecting *expected* reciprocity (e.g., fairness) interact with shocks in complex ways. That is, agents may be afraid that they will be treated unfairly if they receive a bad realization in the Outcome-Only or Effort-Shock treatments. If they believe that they will be unjustly punished if they choose effort equal to the desired effort but receive a negative random number, they may instead choose effort levels lower than desired effort, since high effort is costly. In fact, this may even be an optimal strategy in the presence of shocks. When an agent chooses effort within two levels of desired effort, the marginal gain of an additional unit of effort is only a 20% increase in the principle's perceived probability that at least the desired effort level was given. Thus, agents have incentive at high effort levels to scale back their effort; this saves on rather large costs while minimally decreasing the perceived probability of achieving at least the desired effort. This effect is *exacerbated* if agents are averse to what they view as "unjust" punishment, since the marginal benefit to choosing at least the desired effort is lower when shocks are present.

Neither of these possibilities is mutually exclusive. In fact, they both call for further research on just how and why shocks affect contract choice. While we know that formulating a complete, first-best contract is often not possible when shocks are present, our results suggest that the reciprocity motive does not completely mitigate this problem. Reciprocity does allow for more efficient results than standard contract theory would have us believe, but its effect is reduced by the presence of shocks, whether or not the shocks are observed.

References

Aimone, Jason A., and Daniel Houser. (2011). "Beneficial Betrayal Aversion." *PLoS ONE* 6(3): e17725. doi:10.1371/journal.pone.0017725.

Andreoni, James, and B. Douglas Bernheim. (2009). "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77(5): 1607-1636.

Andreoni, James, William Harbaugh, and Lise Vesterlund. (2003). "The Carrot or the Stick: Rewards, Punishments, and Cooperation." *American Economic Review* 93(3): 893-902.

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy. (2013). "Incentives in Experiments: A Theoretical Analysis." Working Paper.

Baker, George P. (1992). "Incentive Contracts and Performance Measurement." *Journal of Political Economy* 100(3): 598-614.

Baron, Jonathan, and Hershey, John C. (1988). "Outcome bias in decision evaluation." *Journal of Personality and Social Psychology* 54: 569-579.

Bartling, Björn, Ernst Fehr, and Klaus M. Schmidt. (2012). "Screening, Competition, and Job Design: Economic Origins of Good Jobs." *American Economic Review* 102(2): 834-64.

Baumeister, R., Bratslavsky, E., Finkenauer, C., Vohs, K. (2001). "Bad is stronger than good." *Review of General Psychology* 5: 323–370.

Bellemare, Charles & Kroger, Sabine. (2007). "On representative social capital." *European Economic Review* 51(1), 183-202.

Cappelen, Alexander W., James Konow, Erik Ø. Sørensen, and Bertil Tungodden. (2013). "Just Luck: An Experimental Study of Risk-Taking and Fairness" *American Economic Review* 103(4): 1398-1413.

Charness, Gary and Ernan Haruvy. (2002). "Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach." *Games and Economic Behavior* 40: 203-231.

Charness, Gary and Kuhn, Peter. (2011). "Lab labor: What can labor economists learn from the lab?" In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics* 4A: 229-330.

Charness, Gary and Martin Dufwenberg. (2006). "Promises and partnership." *Econometrica* 74: 1579-1601.

Charness, Gary, and D. I. Levine. (2007). "Intention and Stochastic Outcomes: An Experimental study." *Economic Journal* 117: 1051–1072.

Charness, Gary. (2004). “Attribution and Reciprocity in an Experimental Labor Market.” *Journal of Labor Economics* 22(3): 665-688.

Dufwenberg, Martin, and Georg Kirchsteiger. (2004). "A theory of sequential reciprocity." *Games and Economic Behavior* 47: 268-298.

Ericsson, K. Anders and Neil Charness. (1994). “Expert performance: Its structure and acquisition.” *American Psychologist* 49(8): 725-747.

Erkal, Nisvan, Gangadharan, Lata, and Nikos Nikiforakis. (2011). “Relative Earnings and Giving in a Real-Effort Experiment.” *American Economic Review* 101(7): 3330–3348.

Falk, Armin, and Urs Fischbacher. (2006). “A theory of reciprocity.” *Games and Economic Behavior* 54: 293-315.

Fehr, Ernst, and Klaus M. Schmidt. (2007). “Adding a Stick to the Carrot? The Interaction of Bonuses and Fines.” *American Economic Review* 97(2): 177-181.

Fehr, Ernst, and Simon Gächter. (2000). “Fairness and Retaliation: The Economics of Reciprocity.” *Journal of Economic Perspectives* 14(3): 159-181.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger (1997). “Reciprocity as a Contract Enforcement Device: Experimental Evidence.” *Econometrica* 65(4): 833-860.

Fischbacher, Urs. (2007). “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics* 10: 171-178.

Gasper, John T. and Andrew Reeves. (2011). Make It Rain? Retrospection and the Attentive Electorate in the Context of Natural Disasters. *American Journal of Political Science* 55(2): 340-355.

Grossman, Sanford J. and Oliver D. Hart. (1983). “An Analysis of the Principal-Agent Problem.” *Econometrica* 51(1): 7-45.

Gurdal, M., Miller, J., Rustichini, A. (2012). “Why Blame?” Working Paper.

Gürerk1, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. (2006). “The Competitive Advantage of Sanctioning Institutions.” *Science* 312(5770): 108-111.

Harris, Milton and Artur Raviv. (1979). “Optimal Incentive Contracts with Imperfect Information.” *Journal of Economic Theory* 20: 231-259.

Healy, Andrew, Neil Malhotra, and Cecilia Hyunjung Mo. (2010). “Irrelevant events affect voters' evaluations of government performance.” *Proceedings of the National Academy of Sciences* 107(29): 12804-12809.

Holmström, Bengt, and Paul Milgrom. (1991). “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.” *Journal of Law, Economics, & Organization* 7: 24-52.

Holmström, Bengt. (1979). “Moral Hazard and Observability.” *Bell Journal of Economics* 10(1): 74-91.

Houser, Daniel, Erte Xiao, Kevin McCabe, and Vernon Smith. (2008). “When punishment fails: Research on sanctions, intentions, and non-cooperation.” *Games and Economic Behavior* 62: 509-532.

Konow, James. (2000). “Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions.” *American Economics Review* 90(4): 1072-1091.

Konow, James. (2003). “Which Is the Fairest One of All? A Positive Analysis of Justice Theories.” *Journal of Economic Literature* 41(4): 1188-1239.

Laffont, Jean-Jacques and David Martimort. (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton: Princeton University Press.

Marshall, Greg W., and John C. Mowen, (1993). “An Experimental Investigation of the Outcome Bias in Salesperson Performance Evaluations.” *Journal of Personal Selling and Sales Management* 13: 31-47.

Mazzocco, Philip J., Mark D. Alicke, and Teresa L. Davis. (2004). “On the Robustness of Outcome Bias: No Constraint by Prior Culpability.” *Basic and Applied Social Psychology* 26: 131-146.

Milgrom, Paul, and John Roberts. (1992). *Economics, Organization and Management*. Englewood Cliffs: Prentice Hall.

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

Prendergast, Canice. (1999). “The Provision of Incentives in Firms.” *Journal of Economic Literature* 37(1): 7-63.

Rabin, Matthew. (1993). “Incorporating Fairness into Game Theory and Economics.” *American Economic Review* 83(5): 1281-1302.

Rey-Biel, Pedro, Roman Sheremeta, and Neslihan Uler. (2012). “Luck or Effort: Sharing Norms in the US and Europe.” ESI Working Paper.

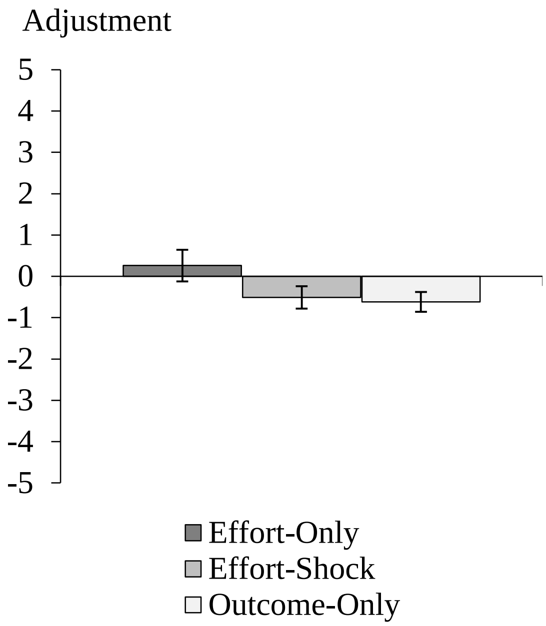
Sefton, Martin, Robert Shupp, and James M. Walker. (2007). "The Effect Of Rewards And Sanctions In Provision Of Public Goods." *Economic Inquiry* 45(4): 671-690.

Shavell, Steven. (1979). “Risk Sharing and Incentives in the Principal and Agent Relationship.” *Bell Journal of Economics* 10(1): 55-73.

Sloof, Randolph, and C. Mirjam Van Praag. (2010). "The effect of noise in a performance measure on work motivation: A real effort laboratory experiment." *Labour Economics* 17: 751-765.

Xiao, Erte and Howard Kunreuther. (2012). "Punishment and Cooperation in Stochastic Prisoner’s Dilemma Game." NBER Working Paper.

Figure 1: Average Adjustment.



Note: the error bars represent the standard error of the mean.

Figure 2: Distribution of Adjustment.

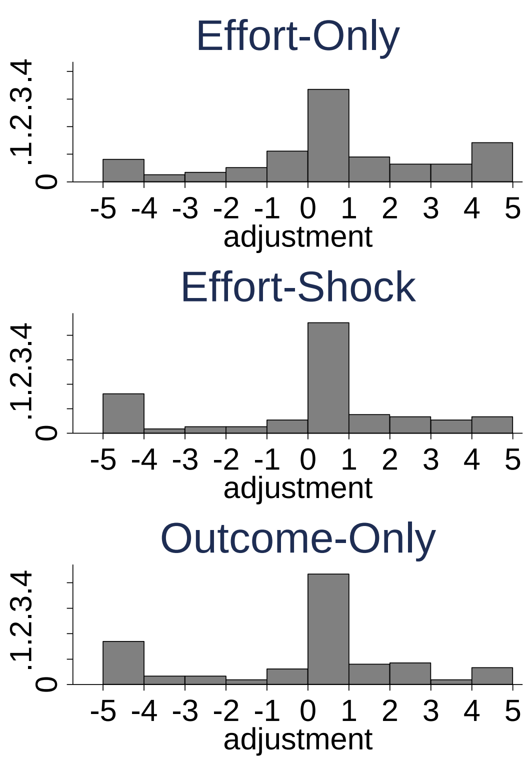
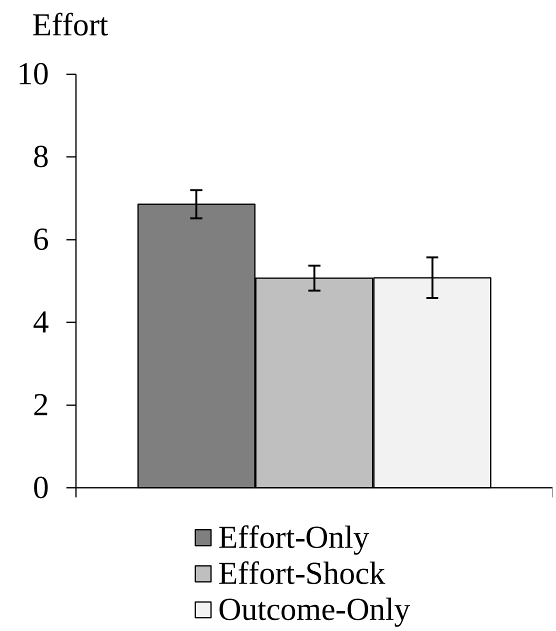


Figure 3: Average Effort.



Note: the error bars represent the standard error of the mean.

Figure 4: Distribution of Effort.

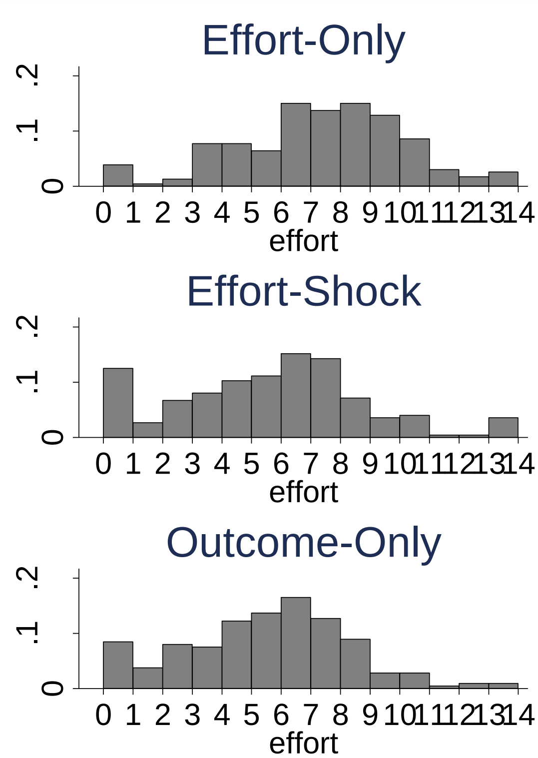


Figure 5: Distribution of Effort – Desired Effort.

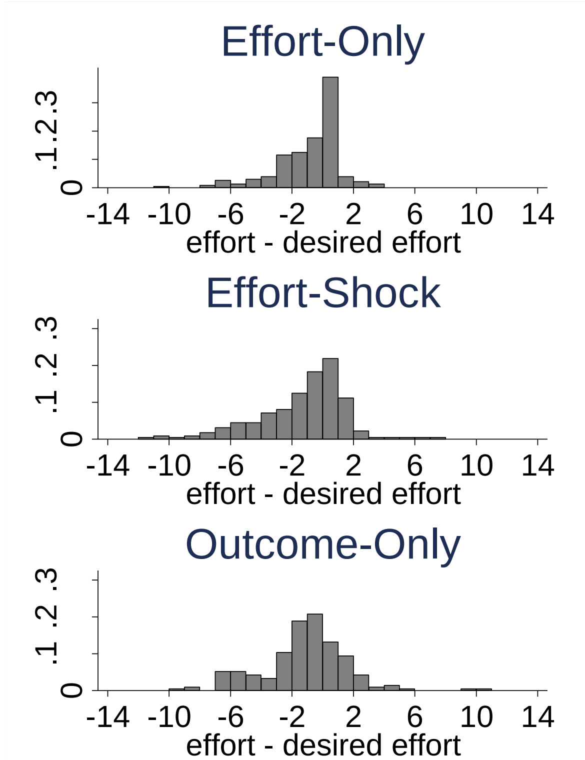
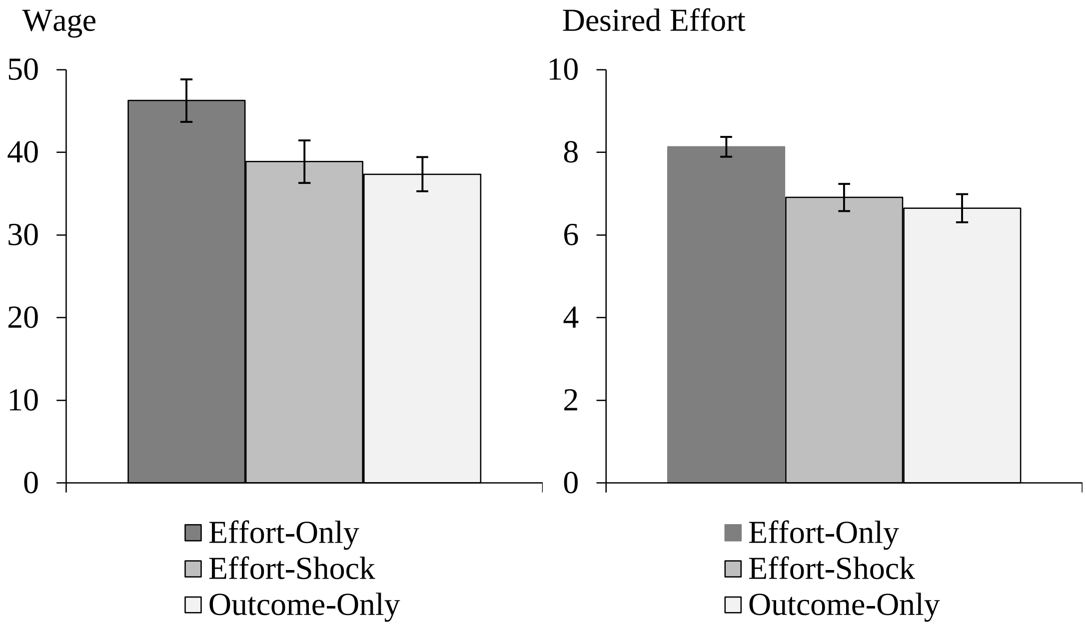


Figure 6: Average Wage and Desired Effort.



Note: the error bars represent the standard error of the mean.

Figure 7: Distribution of Wage and Desired Effort.

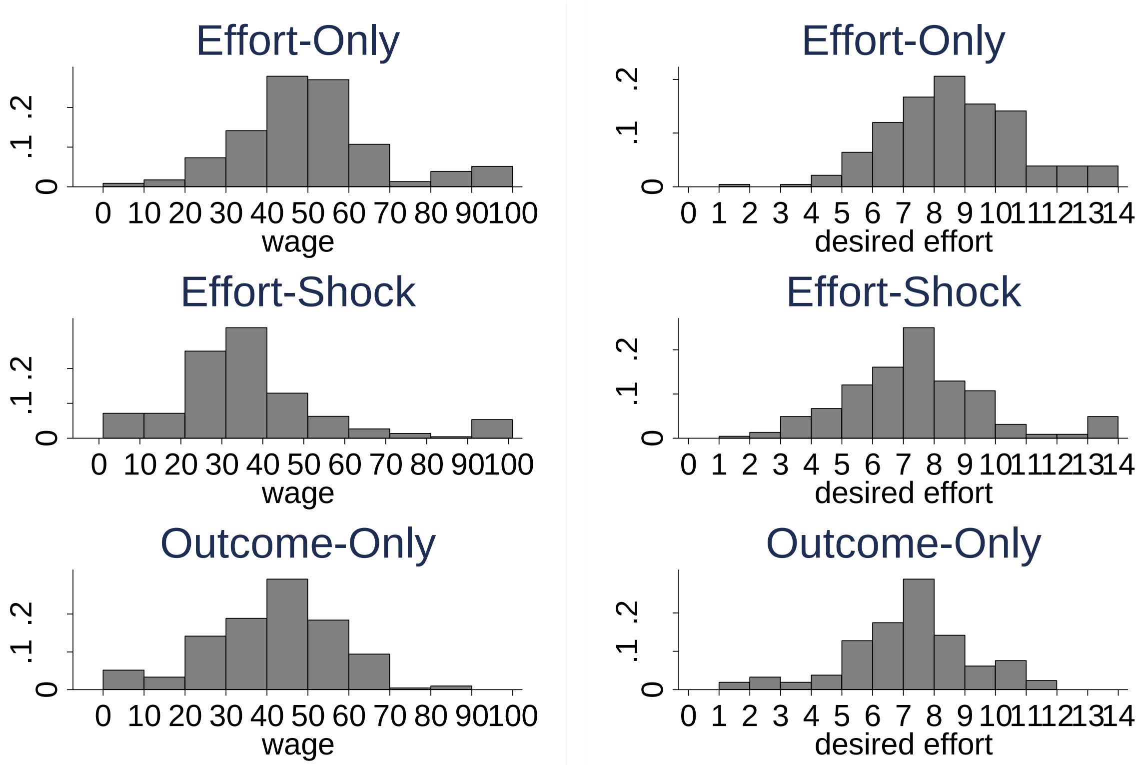


Table 1: The Average Statistics: IR/IC Contracts versus All Contracts.

Treatment	Wage	Desired Effort	Effort	Outcome	Adjustment	Principal's Payoff	Agent's Payoff	Total Welfare
				Contracts Satisfying IR/IC				
Effort-Only	46.27	8.13	6.86	6.86	0.26	20.39	18.91	39.30
	(2.57)	(0.24)	(0.34)	(0.34)	(0.38)	(2.89)	(1.68)	(1.48)
Effort-Shock	38.87	6.91	5.07	4.98	-0.51	9.17	20.27	29.44
	(2.56)	(0.33)	(0.30)	(0.33)	(0.27)	(3.75)	(2.06)	(2.19)
Outcome-Only	37.36	6.65	5.08	5.12	-0.62	12.09	19.50	31.58
	(2.06)	(0.34)	(0.49)	(0.48)	(0.24)	(3.65)	(2.10)	(2.33)
				All Contracts				
Effort-Only	41.14	8.95	6.40	6.40	0.14	20.91	15.71	36.62
	(3.22)	(0.30)	(0.43)	(0.43)	(0.34)	(3.11)	(1.88)	(1.91)
Effort-Shock	33.45	7.63	4.69	4.62	-0.18	11.23	16.40	27.64
	(2.98)	(0.33)	(0.34)	(0.32)	(0.22)	(3.46)	(2.49)	(1.90)
Outcome-Only	33.85	7.63	4.69	4.75	-0.50	12.04	17.63	29.67
	(2.28)	(0.25)	(0.41)	(0.38)	(0.13)	(2.85)	(1.89)	(2.11)

Standard errors in parenthesis (based on 9 independent observations).

Table 2: Panel Models of Adjustments.

Specification	(1)	(2)	(3)	(4)	(5)
Treatments	*Effort-Only*	*Effort-Shock*	*Outcome-Only*	*Effort-Shock*	*Effort-Shock*
Dependent variable	*adjustment*	*adjustment*	*adjustment*	*adjustment*	*adjustment*
Wage	0.01*	0.01***	0.02	0.01***	0.01***
[wage]	(0.01)	(0.00)	(0.02)	(0.00)	(0.00)
effort – desired effort	0.78***	0.45***		0.46***	
[effort gap]	(0.12)	(0.07)		(0.08)	
outcome – desired effort			0.44***		0.44***
[outcome gap]			(0.10)		(0.08)
Shock				0.35**	
[random number]				(0.14)	
period	-1.03**	-0.56	-0.14	-0.63	-0.62
[inverse period]	(0.41)	(0.44)	(0.55)	(0.44)	(0.44)
Constant	1.02	0.11	-0.54	0.15	0.13
[constant term]	(0.67)	(0.30)	(0.82)	(0.28)	(0.30)
Observations	233	224	212	224	224
Clusters	9	9	9	9	9
Overall R-squared	0.33	0.23	0.20	0.27	0.27

* significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors in parenthesis are clustered at the group level.

Table 3: Panel Models of Effort.

Specification	(1)	(2)	(3)
Treatments	*Effort-Only*	*Effort-Shock*	*Outcome-Only*
Dependent variable	*effort*	*effort*	*effort*
Wage	0.04***	0.03*	0.04*
[wage]	(0.01)	(0.02)	(0.02)
desired effort	0.61***	0.41**	0.12
[desired effort]	(0.11)	(0.16)	(0.17)
Period	0.01	0.82	0.03
[inverse period]	(0.55)	(0.76)	(0.38)
Constant	0.01	0.73	2.92***
[constant term]	(0.54)	(0.63)	(0.71)
Observations	233	224	212
Clusters	9	9	9
Overall R-squared	0.49	0.27	0.16

* significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors in parenthesis are clustered at the group level.

Appendix A (Not for Publication): Instructions for the Effort-Shock Treatment

INSTRUCTIONS

This is an experiment in the economics of strategic decision making. Various research agencies have provided funds for this research. The instructions are simple. If you follow them closely and make appropriate decisions, you can earn an appreciable amount of money.

The currency used in the experiment is francs. Francs will be converted to U.S. Dollars at a rate of **_10_** francs to **_1_** dollar. You have already received a **$20.00** participation fee (this includes your show-up fee of $7.00). Your earnings from the experiment will be incorporated into your participation fee. At the end of today's experiment, you will be paid in private and in cash. There are **24** participants are in today's experiment.

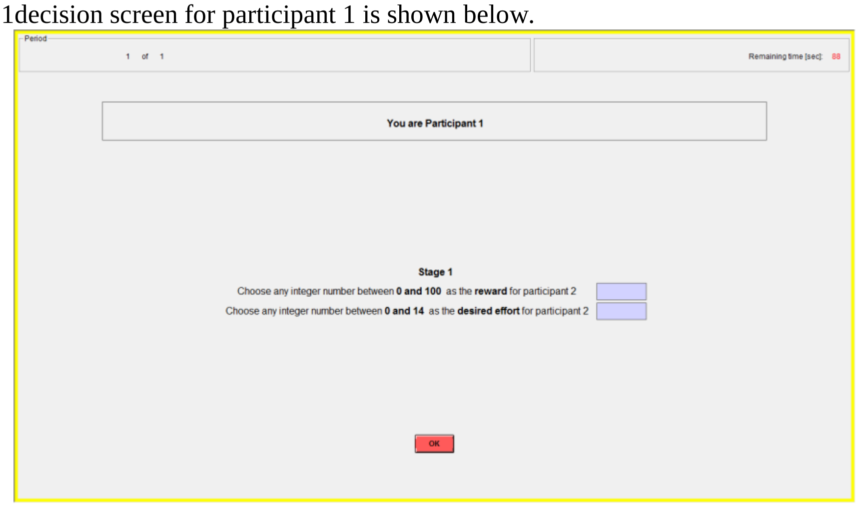
It is very important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

YOUR ROLE ASSIGNMENT

The experiment consists of **10 decision-making periods**. Each period, you will be randomly and anonymously placed into a group which consists of two participants: **participant 1** and **participant 2**. At the beginning of the first period you will be randomly assigned either as participant 1 or participant 2. You will **remain in the same role assignment** throughout the entire experiment. So, if you are assigned as participant 2 then you will stay as participant 2 throughout the entire experiment. Each consecutive period you will be **randomly re-grouped** with another participant of opposite assignment. So, if you are participant 2, each period you will be randomly re-grouped with another participant 1.

STAGE 1

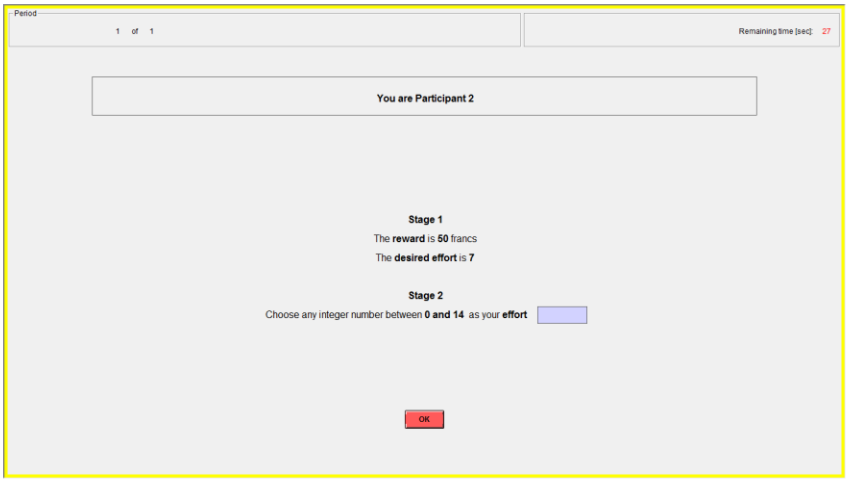
Each period will proceed in three stages. In Stage 1, **participant 1** will choose a **reward** (any integer number between **0 and 100**) and a **desired effort** (any integer number between **0 and 14**) for participant 2. An example of Stage 1decision screen for participant 1 is shown below.



[Stage 1 decision screen]

STAGE 2

The computer will display to participant 2 the reward and the desired effort chosen by participant 1. Then in Stage 2, **participant 2** will choose an **effort** level (any integer number between **0 and 14**). An example of Stage 2 decision screen for participant 2 is shown below.



[Stage 2 decision screen]

For each **effort** level chosen by participant 2 there is an associated **cost of effort**. The cost of effort can be found in the following table:

Effort	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
cost of effort	0	1	2	5	8	13	18	25	32	41	50	61	72	85	98

Note that as effort rises from 0 to 14, costs rise exponentially. The cost of effort can be also calculated using the following formula (and rounding the number to the nearest highest integer):

$$\text{participant 2's cost of effort} = \frac{(\text{effort})^2}{2}$$

STAGE 3

After participant 2 chooses the effort level, the computer will add to effort a **random number** to determine the **performance** of participant 2:

participant 2's performance = effort + random number

The random number chosen by the computer can take a value of **-2, -1, 0, 1, or 2**. Each number is equally likely to be drawn. After the computer makes a draw of a random number, it will display to participant 1 the **performance** of participant 2, as well as the **effort** chosen by participant 2 and the **random number** chosen by the computer. Then in the third stage, **participants 1** will choose an **adjustment** level. The adjustment level can be any number, multiple of 10, between **-50 and 50**. An example of Stage 3 decision screen for participant 1 is shown below.

[Stage 3 decision screen]

For each **adjustment** level chosen by participant 1 there is an associated **cost of adjustment**. The cost of adjustment can be found in the following table:

adjustment	-50	-40	-30	-20	-10	0	10	20	30	40	50
cost of adjustment	5	4	3	2	1	0	1	2	3	4	5

EARNINGS OF PARTICIPANTS 1 AND 2

The earnings of participant 1 depend on the **reward** chosen by participant 1 in the first stage, the **performance** of participant 2 in the second stage and the **adjustment** chosen by participant 1 in the third stage. Specifically, the participant 1's earnings are calculated by the following formula:

participant 1's earnings = 10*(performance) – (reward) – (cost of adjustment) =
= 10*(effort + random number) – (reward) – (cost of adjustment)

Note that higher participant 2's effort implies higher participant 2's performance, and thus higher participant 1's earnings. On the other hand, the higher reward or the higher cost of adjustment implies lower participant 1's earnings.

The earnings of participant 2 depend on the **reward** chosen by participant 1 in the first stage, the **effort** chosen by participant 2 in the second stage and the **adjustment** chosen by participant 1 in the third stage. Specifically, the participant 2's earnings are calculated by the following formula:

participant 2's earnings = (reward) – (cost of effort) + (adjustment)

Note that higher reward chosen by participant 1 implies higher participant 2's earnings. On the other hand, the higher effort implies lower participant 2's earnings. If participant 1 choses a positive adjustment level for participant 2 then participant 2's earnings increase by that adjustment level. However, if participant 1 choses a negative adjustment level then participant 2's earnings decrease by that adjustment level.

Example 1

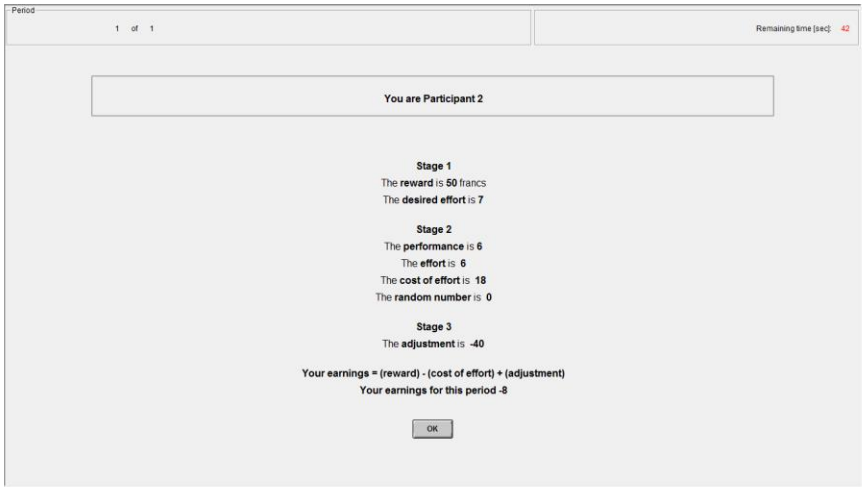
Assume the following scenario. In the first stage, participant 1 chooses a reward of 50 and a desired effort of 7. In the second stage, participant 2 chooses an effort of 6. Then the computer randomly selects 2 as a random number, so the performance of participant 2 is 8 (6+2). Then the computer displays to participant 1 that participant 2's performance is 8, participant 2's effort is 6 and the random number chosen by the computer is 2. After observing this information, in the third stage, participant 1 chooses an adjustment of -40. Therefore, participant 1's earnings = 10*8 – 50 – 4 = 26, since participant 2's performance is 8, the reward is 50, and the cost of adjustment of -40 is 4. Finally, participant 2's earnings = 50 – 18 – 40 = -8, since the reward is 50, the cost of effort of 6 is 18, and the adjustment is -40.

Example 2

Assume the following scenario. In the first stage, participant 1 chooses a reward of 40 and a desired effort of 6. In the second stage, participant 2 chooses an effort of 9. Then the computer randomly selects -2 as a random number, so the performance of participant 2 is 7 (9-2). Then the computer displays to participant 1 that participant 2's performance is 7, participant 2's effort is 9 and the random number chosen by the computer is -2. After observing this information, in the third stage, participant 1 chooses an adjustment of 30. Therefore, participant 1's earnings = 10*7 – 40 – 3 = 27, since participant 2's performance is 7, the reward is 40, and the cost of adjustment of 30 is 3. Finally, participant 2's earnings = 40 – 41 + 30 = 29, since the reward is 40, the cost of effort of 9 is 41, and the adjustment is 30.

END OF THE PERIOD

At the end of each period, the computer will calculate individual earnings and display to both participants the following information: the reward chosen by participant 1, the desired effort chosen by participant 1, the performance of participant 2, the effort chosen by participant 2, the random number chosen by the computer, the adjustment chosen by participant 1, as well as individual earnings for that period. An example of the outcome screen is shown below.



[Outcome screen]

Once your earnings are displayed on the outcome screen as shown below you should record your earnings for the period on your **Personal Record Sheet** under the appropriate heading.

IMPORTANT NOTES

Each period, you will be randomly and anonymously placed into a group which consists of two participants: **participant 1** and **participant 2**. At the beginning of the first period you will be randomly assigned either as participant 1 or participant 2. You will **remain in the same role assignment** throughout the entire experiment. So, if you are assigned as participant 2 then you will stay as participant 2 throughout the entire experiment. Each consecutive period you will be **randomly re-grouped** with the other participant of opposite assignment. So, if you are participant 2, each period you will be randomly re-grouped with another participant 1.

Each period will proceed in three stages. In Stage 1, **participant 1** will choose a **reward** and a **desired effort** for participant 2. The computer will display to participant 2 the reward and the desired effort chosen by participant 1. Then in Stage 2, **participant 2** will choose an **effort** level. For each **effort** level chosen by participant 2 there is an associated **cost of effort**. After participant 2 chooses the effort level, the computer will add to effort a **random number** to determine the **performance** of participant 2. Then the computer will display to participant 1 the **performance** of participant 2, as well as the **effort** chosen by participant 2 and the **random number** chosen by the computer. Then in Stage 3, **participant 1** will choose an **adjustment** level. Finally, at the end of each period, the computer will calculate individual earnings and display both participants all relevant information.

Remember you have already received a **$20.00** participation fee. In the experiment, depending on a period, you may receive either positive or negative earnings. At the end of the experiment we will randomly select **1 out of 10** periods for actual payment and convert them to a U.S. dollar payment. If the earnings are negative, we will subtract them from your total earnings. If the earnings are positive, we will add them to your total earnings.

Are there any questions?

Cost of Effort Table

For each **effort** level chosen by participant 2 there is an associated **cost of effort**.

effort	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
cost of effort	0	1	2	5	8	13	18	25	32	41	50	61	72	85	98

Cost of Adjustment Table

For each **adjustment** level chosen by participant 1 there is an associated **cost of adjustment**.

adjustment	-50	-40	-30	-20	-10	0	10	20	30	40	50
cost of adjustment	5	4	3	2	1	0	1	2	3	4	5

Appendix B (Not for Publication): Non-IR/IC Contracts

The analysis in Section 3 is based on contracts that satisfy individual rationality and incentive compatibility constraints (IR/IC), i.e. contracts in which both the principal's and the agent's payoffs are non-negative, conditional on the contract being exactly fulfilled. Often researchers deliberately restrict subject's strategy space so that they can choose only contracts which satisfy IR/IC (i.e., Bartling et al., 2012). In our experiment, we chose not to put such restrictions on the principal's decisions, because some ex-ante non-IR/IC contracts (w,$\underline{e}$) may actually give positive payouts for both the principal and the agent ex-post, given a certain level of adjustment a.

Before we examine non-IR/IC contracts, we emphasize two facts. First, the majority of contracts observed in our experiment satisfy IR/IC (62%). Moreover, the number of contracts satisfying IR/IC is very similar across the three treatments (64% in Effort-Only, 59% in Outcome-Only and 62% in Effort-Shock). Second, practically all results (i.e., Results 1-6) reported in Section 3 hold when we combine IR/IC and non-IR/IC contracts together.[20] This can be easily seen be by comparing the top and the bottom part of Table 1. The only reason we chose to focus mainly on IR/IC contracts is that these contracts are easy to interpret economically and they are also most likely to be employed in practice.

Examining the non-IR/IC contracts, we find that 36% of contracts in the Effort-Only, 41% in the Outcome-Only and 38% in the Effort-Shock treatment can be characterized as non-IR/IC contracts. The vast majority of the non-IR/IC contracts (91%, 744 out of 822) are the ones where, conditional on contract being exactly fulfilled, the agent is expected to receive a negative payoff (i.e., $w - c(\underline{e}) < 0$). We refer to these contracts as the "stingy" contracts, since they do not satisfy the incentive compatibility requirements for the agent. The rest of the non-IR/IC contracts (9%, 78 out of 822) are the ones where the principal is expected to receive a negative payoff (i.e., $10\underline{e} - w < 0$). We refer to these contacts as the "generous" contacts, since in such contracts the principal offers very generous wages relatively to the desired effort, although they are not

[20] There are three results which do not hold exactly but come very close to statistical significance. For Result 1, the conditional adjustment is marginally higher in Effort-Only versus Outcome-Only (p-value = 0.08). For Result 3, there is still a higher probability that effort is greater than or equal to desired effort in the Effort-Only treatment than in the Outcome-Only treatment, but with weaker statistical significance (p-value = 0.14). For Result 4, there is still a greater wage and desired effort in the Effort-Only treatment than in Outcome-Only, but with weaker statistical significance (p-value = 0.15).

individually rational. Table B1 provides the summary statistics across all three treatments, using "stingy" contracts (top part) and "generous" contracts (bottom part).

Examining first the 78 generous contracts where the principal is expected to make a negative payoff (the bottom panel of Table B1), we find that principals make negative payoffs in all three treatments. Due to the small number of such contracts there are only 5 independent observations in the Effort-Only treatment, 4 in the Effort-Shock treatment, and 6 in the Outcome-Only treatment. Although a full statistical analysis is not appropriate with such a small (and very noisy) number of independent observations, it appears that when the principals offer "generous" contracts, the agents reciprocate. The effort is about 2 units higher than the desired effort, which is in sharp contrast to the IR/IC contacts in Table 1. However, this reciprocation by agents is not nearly enough to compensate for very generous wage offers by principals. As a consequence, principals make negative payoffs, while agents make positive and very high payoffs. It is important to emphasize that such behavior by principals is unlikely to be caused by mistakes. The 78 generous contracts are offered by 52 independent subjects. Moreover 34 of these contacts are offered in the last five periods of the experiment. A possible explanation for generous contracts is that principals use such contracts hoping that agents will show significant reciprocity. While we cannot rule out the possibility that generous contracts are caused by mistakes (although it is unlikely), we can rule out the possibility that these contracts offer "efficiency wages" – aimed at keeping agents happy in the long-run – since relationships are one-shot.

When examining the 744 stingy contracts where the agent is expected to make a negative payoff (the top panel of Table B1), we find that in all three treatments agents make significantly positive payoffs (based on the Wilcoxon signed-rank test, separately for each treatment, all p-values are less than 0.01). This is mainly because the effort is about 5-6 units below the desired effort (which is in sharp contrast to the generous contracts in Table B1 and the IR/IC contacts in Table 1). Interestingly, on average, principals do not punish such behavior. One reason for this is that principals make very substantial payoffs even when agents do not fully fulfill the terms of the stingy contract. In fact, the principals offering stingy contracts on average receive very similar payoffs than the principals offering IR/IC contracts (based on the Wilcoxon signed-rank test, separately for each treatment, all p-values are greater than 0.20). Agents, on the other hand, receive significantly lower payoffs under the stingy contracts than under the IR/IC contracts (based on the Wilcoxon signed-rank test, separately for each treatment, all p-values are less than 0.02). These two findings can explain why the majority of contracts satisfy IR/IC, yet some principals still choose to offer stingy contracts. On the one hand, principals should be indifferent between offering the IR/IC and stingy contracts, since the expected payoff is not different between the two. So, we should observe both types of the contacts. On the other hand, IR/IC contracts provide significantly higher payoffs to the agents, and thus benevolent principals should choose the IR/IC contracts more often than the stingy contracts.

Table B1: The Average Statistics: Non-IR/IC Contracts.

Treatment	Wage	Desired Effort	Effort	Outcome	Adjustment	Principal's Payoff	Agent's Payoff	Total Welfare
				Stingy Contracts				
Effort-Only	30.31	10.68	5.35	5.35	-0.08	21.23	9.42	30.64
	(3.94)	(0.51)	(0.75)	(0.75)	(0.41)	(4.96)	(2.37)	(3.45)
Effort-Shock	23.00	9.12	3.83	3.66	0.22	12.44	9.49	21.93
	(3.92)	(0.47)	(0.50)	(0.49)	(0.31)	(4.17)	(3.05)	(3.04)
Outcome-Only	21.21	10.38	4.00	4.09	-0.50	18.23	7.62	25.85
	(4.03)	(0.51)	(0.56)	(0.53)	(0.30)	(3.95)	(3.02)	(3.24)
				Generous Contacts				
Effort-Only	54.00	2.20	4.00	4.00	-2.00	-16.00	42.00	26.00
	(8.27)	(0.80)	(0.95)	(0.95)	(1.22)	(11.53)	(9.82)	(4.16)
Effort-Shock	44.88	1.88	3.44	4.19	-0.94	-4.06	36.50	32.44
	(6.71)	(0.72)	(0.86)	(1.12)	(0.60)	(8.06)	(5.30)	(9.80)
Outcome-Only	57.47	2.96	5.44	5.61	0.44	-3.07	35.38	32.31
	(6.46)	(0.33)	(0.88)	(1.02)	(0.60)	(13.62)	(11.62)	(3.83)

Standard errors in parenthesis (based on 9 independent observations). In the bottom panel of the table there are only 5 independent observations in the Effort-Only treatment, 4 in the Effort-Shock treatment, and 6 in the Outcome-Only treatment.