Principal-Agent Settings with Random Shocks

Jared Rubin and Roman Sheremeta

5 February 2015

Principal-Agent Settings with Random Shocks

Jared Rubin [a,b] and Roman Sheremeta [b,c,*]

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

February 4, 2015

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. Therefore, our findings demonstrate that random shocks place an important bound on the ability of gift exchange to overcome principal-agent problems.

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

---

* Corresponding author: Roman Sheremeta, rshereme@gmail.com and rms246@case.edu

We thank three anonymous referees and the Editor of this journal for their valuable suggestions. We have benefitted from the helpful comments of Gary Charness, Brice Corgnet, Ron Harstad, Roberto Hernán-González, Alex Imas, Charlie Plott, David Rojo-Arjona, John List, three anonymous referees, and seminar participants at Chapman University and Case Western Reserve University, as well as participants at the North-American Economic Science Association Meeting, the 2013 International Foundation for Research in Experimental Economics Conference at Chapman University, and the 2014 ASSA Annual Meeting. 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 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

[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).

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 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*

[4] There is a wealth of experimental evidence that both positive 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.

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.

What can explain our findings? Why should the introduction of a shock reduce wages and effort if the shock is perfectly observable? To address this question, we outline a model of reciprocity (in the context of our experiment) where subjects reciprocate based on either the *effort* or *outcome* of the previous game play.[5] If wages and effort are solely encouraged by effort-based reciprocity, 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. On the other hand, if wages and effort are solely encouraged by outcome-based reciprocity, there should be no difference between the treatment where shocks are perfectly observed and the treatment where shocks are not observed, since the mapping from effort to expected outcome (and reciprocity) is the same in both cases. Moreover, the effort exerted in these two treatments should be

[5] We consider only these cases where subjects reciprocate based solely on effort or outcome. In reality, it is likely that the reciprocity motive is some combination of the two. The implication is that the actual outcome should be somewhere in the middle of the two proposed outcomes.

lower than in the baseline treatment, since it is more costly to “make up” for a bad shock than it is to scale back effort for a “good shock” (the cost curve is increasing and convex).

We find evidence in favor of subjects exhibiting outcome-based reciprocity. In the treatment where principals observe both the agent’s effort and the shock, principals do indeed vary their adjustments based on the shock, which is outside the control of the agent. This suggests that reciprocity is in part influenced by the outcome the principal receives, even if the principal knows that part of this outcome was influenced by luck. Moreover, as the outcome-based reciprocity hypothesis suggests, wages and effort are significantly lower in treatments where shocks are perfectly observed relative to the baseline, and 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 a number of important implications. First, our results provide evidence that the reciprocity motive in the principal-agent settings 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). In this regard, we contribute to the literature studying how individual behavior is impacted by intentions and outcomes (Falk and Fischbacher, 2006). We show that this result has significant welfare implications: welfare-enhancing effort is lower in the presence of shocks, *even when* the shocks are perfectly observable. Second, our results contribute to the large literature on gift exchange. Charness and Kuhn (2011) review the experimental evidence on gift exchange, concluding that gift exchange is a robust phenomenon in that higher wages lead to higher effort. Our study contributes to this literature by demonstrating that the existence of random shocks is an important boundary condition of gift exchange. To this end, our study adds to an important literature, highlighted by List (2007), examining how the introduction of realistic elements and institutions into gift exchange settings impacts individual behavior (e.g., Gneezy and List, 2006; Falk and Kosfeld, 2006; Charness and Gneezy, 2008). Gneezy and List (2006), for example, show that positive reciprocity effects detected in laboratory gift exchange experiments can wear off quickly in the field. Similarly, Falk and Kosfeld (2006) show that reciprocity declines when principals try to control agents’ performance. On the other hand, Charness and Gneezy (2008) show that agents are more reciprocal when anonymity is reduced. Our results add yet one more realistic element, showing 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 (as is usually the case in the real world).

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 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 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 $\pi^A = w - c(e) + 10a$.[8] The range of payoffs in any one period can vary substantially for both players, ranging from -105 to 160 for the principal and -148 to 150 for the agent.[9]

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.[10]

[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).

[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] In the treatments with shocks, the entire burden of the risk is placed on the principal, as the shock is realized after the agent chooses effort and the shock directly enters the principal's payoff function.

[9] The principal can receive a payoff of up to 140 in the treatment without shocks. In the experiment, principals' single period payouts ranged from -95 to 150 and agents' single period payouts ranged from -121 to 124.

[10] 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 (2015), 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.

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., $a = 5$), providing a total welfare gain of 95 (10*10 – 50 + 50 – 5).[11]

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 (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.[12] 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).[13] 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).

---

[11] This does mean that nearly half of the welfare-maximizing contract comes from the bonus, while we are primarily concerned with the welfare implications of the effort chosen by the agent. Yet, since the bonus is costly and is chosen in the final stage of a one-shot game, there is no opportunity for the principal to receive any future reciprocity from the agent. Hence, we expect that most of the surplus gain will come from effort and not from the bonus. We indeed find this; Figure 2 indicates that in no treatment did more than 10% of principals give the highest possible adjustment of 5.

[12] 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 4. Subjects were not told that they were split into three groups of eight.

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

At the end of each experiment, 1 out of 10 periods were randomly selected for payment.[14] 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.[15]

3. Predictions

Before proceeding to the results of the experiment, we provide intuition and predictions for how subjects might act under different experimental settings. To this end, we verbally discuss a model focusing on the reciprocity motive.[16] After all, the experiment is centered on a double gift exchange – with agents gifting principals with effort and principals gifting agents with adjustments – and reciprocity is a key motivation of gift exchange.

It is not obvious what might motivate people to reciprocate in the context of our experiment. The literature provides some insights, as numerous studies have shown that subjects often reciprocate based on both *effort* and *outcome*. 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. 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; Cappelen et al., 2013; Gurdal et al., 2013; Rey-Biel et al., 2015). Hence, in this section, we present intuition for how the reciprocity motive might affects individual actions in the context of our experiment under two conditions: (1) principals reward agents based solely on effort, and (2) principals reward agents based on the outcome of the agent's actions (effort plus shock).[17] In reality, it is

---

[14] 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 randomly selected period is the only mechanism (under a wide array of assumptions) that mitigates this inter-period problem that could otherwise cause some loss of control for the experimenter.

[15] 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.

[16] We drafted a formal model, and we found that the equilibrium outcomes are dependent on how reciprocity is modeled. The two extreme forms of reciprocity studied in this section – where reciprocity is either based solely on effort or on outcome – provide the same results in the formal model as they are described in this section. Since we did not design the experiment to extract the shape of the reciprocity function – meaning that we cannot derive any meaningful testable predictions with respect to its shape – and the "extreme" results are straight-forward to discuss verbally anyhow, we have only included a verbal discussion of the model in order to facilitate the reader's intuition for the results that we find in our analysis.

[17] Focusing on the case where principals reward agents based on the *fairness* of the agent's actions (Rabin 1993; Fehr and Gächter, 2000) gives qualitatively similar predictions. This is true, specifically, under the following fairness principle: if the agent anticipates a positive shock (with some probability), it is "fair" for the agent to split some of the surplus with the principal in the form of lower effort (as long as the outcome is also not lower), while if

likely that subjects will reciprocate based on some combination of effort and outcome, as the above cited papers suggest. Yet, focusing on the two extreme cases allow us to shed some light on which mechanism is more important in determining outcomes.

In the third stage, principals are able to show reciprocity for high (low) effort or outcome with a positive (negative) and costly adjustment. We assume that the principals' reciprocity motive is a relative one; that is, it is a function of the agent's effort (or outcome) relative to the desired effort. If principals reward agents based solely on their effort and not the outcome, this means that the adjustment does not vary with the size of the shock, since the shock is outside the agent's control. On the other hand, if principals reward agents based on the outcome of their actions, then principals account for the fact that there is a component of the outcome that is beyond the agent's control, and the adjustment will vary with the shock. We also assume that agents show reciprocity as an increasing function of the wage that principals give them in the first stage. Specifically, we provide predictions for two scenarios:

Hypothesis 1: Principals reciprocate based on *effort*, meaning that principals set their adjustments based on the effort given by agents relative to desired effort, not the outcome.

Hypothesis 2: Principals reciprocate based on *outcome*, meaning that principals set their adjustments based on the outcome (effort plus shock) relative to desired effort.

In our experiment, three treatments are considered: Effort-Only (where there is no shock), Effort-Shock (where the principal sees the effort and the shock before choosing an adjustment), and Outcome-Only (where the principal sees the outcome, but not the effort or shock value, before choosing an adjustment). As we show below, the two hypotheses provide different testable predictions for the Effort-Only and Effort-Shock treatments. Hence, comparing these two treatments allows us to falsify at least one of the hypotheses. We therefore begin by discussing testable predictions for only the Effort-Only and Effort-Shock treatments.

Consider Hypothesis 1, where principals reciprocate based solely on effort (relative to desired effort), and agents know that principals reciprocate in this manner. In this case, the adjustment given by the principal should not vary as the shock varies. Hence, the existence of a shock should not affect the agent's effort in equilibrium (conditional on wage and desired effort) in the Effort-Shock treatment relative to the Effort-Only treatment, since principals can perfectly observe the agent's effort and the shock. There should also be no difference in the principal's wage and desired effort in these two

---

the agent anticipates a negative shock (with some probability), it is "fair" for the agent to make up for some of the lost surplus with extra effort (since the principal is the residual claimant of the lost surplus).

treatments, as agents should be expected to react the same to these two choices in both treatments. This logic is summarized in Prediction 1.

Prediction 1: If principals reciprocate based solely on *effort* (Hypothesis 1), then adjustments should not vary across shock levels (conditional on effort) in the Effort-Shock treatment, and there should be no difference in any of the subject's choices (wage, desired effort, effort, adjustment) in the Effort-Only and Effort-Shock treatments.

Next, consider Hypothesis 2, where principals reciprocate based on the outcome. In this case, the principal's adjustment is increasing in the shock in the Effort-Shock treatment – as the shock increases, the outcome increases, which motivates greater reciprocation. When agents choose their effort level, they know that there is a 0.2 probability that they will receive each of the shocks from the set {-2, -1, 0, 1, 2}. Since effort plays a smaller role in determining the adjustment in the Effort-Shock treatment than in the Effort-Only treatment – the shock plays a role in determining the adjustment in the former but not the latter – the net marginal return of higher effort is lower in the Effort-Shock treatment.[18] Since effort is costly and increasing in a convex manner, the incentive to "make up" for lost effort in the negative shock case is less than the incentive to reduce effort created by a positive shock (due to the convex nature of the cost of effort curve), and the weighted effort is therefore lower in the Effort-Shock treatment than in the Effort-Only treatment. As a result, principals give a lower wage and ask for less desired effort in the Effort-Shock treatment, since their "gift" (i.e., wage), is less effective at inducing effort. This logic is summarized in Prediction 2.

Prediction 2: If principals reciprocate based on the *outcome* (Hypothesis 2), then adjustments should be increasing in the shock in the Effort-Shock treatment, and the average effort, wage, and desired effort are lower in the Effort-Shock treatment than in the Effort-Only treatment.

These two predictions allow us to falsify either Hypothesis 1 or 2 (or both). Of course, we are also interested in how subjects behave in the Outcome-Only treatment; indeed, this is the treatment that is

[18] To see this, assume that agents place equal weight on all five of the possible shocks and act according to the weighted sum of their different actions. One-fifth of the time there is a zero shock, and the adjustment given by principals (conditional on effort, desired effort, and wage) should be the same as in the Effort-Only treatment, since effort equals outcome. With 0.4 probability there is a positive shock. In this case, the agents want to give less effort than when there is a zero shock, but only to the extent that the outcome is not too small. Finally, with 0.4 probability there is a negative shock. Here, agents may want to give more than in the zero shock case in order to "make up" for the negative shock. However, the incentive to do so is partially mitigated by the fact that this involves a greater cost of effort, which is increasing at an increasing rate.

most similar to real world principal-agent settings. In this treatment, principals do not observe effort, only the outcome. This gives agents the opportunity to "hide behind randomness" in stage 2 (Andreoni and Bernheim, 2009; Aimone and Hauser, 2011), acting selfishly when they can ascribe their actions to chance.

We focus here on predictions for the Outcome-Only treatment under Hypothesis 2.[19] In stage 3, principals can only see the outcome, not the effort or shock. Therefore, the principal should show greater reciprocity as the outcome increases (relative to desired effort) and the adjustment should be the same as in the Effort-Shock treatment. Given this logic, the agent's choice of effort in stage 2 is exactly the same in the Outcome-Only and Effort-Shock treatments. In both treatments, the agent does not know the shock value when choosing effort, and the only thing that matters to the principal when choosing the adjustment is the outcome. Hence, the decision-making calculus is the same for the agent in both treatments. From Prediction 2, this entails that the average effort as well as the average wage and desired effort are lower in the Outcome-Only treatment than in the Effort-Only treatment. This logic is summarized in Prediction 3.

Prediction 3: If principals reciprocate based on the *outcome* (Hypothesis 2), then adjustments, average effort, wage, and desired effort should be the same in the Outcome-Only and Effort-Shock treatments, while the average effort, wage, and desired effort are lower in the Outcome-Only treatment than in the Effort-Only treatment.

4. Results

We observed 2160 contracts in our experiment.[20] Table 1 provides the summary statistics across all three treatments. 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

[19] We find evidence contrary to Prediction 1 in the following section, suggesting that principals do not reciprocate based solely on effort. Hence, we do not discuss predictions of the Outcome-Only treatment under Hypothesis 1. Solving for how subjects act under this hypothesis is not trivial and the direction of the comparative statics (vis-à-vis other treatments) depend on the level of the choices in the other treatments.

[20] Out of 2160 contracts, 1338 (62%) can be classified as individually rational and incentive compatible (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, IR/IC contracts $(w,\underline{e})$ satisfy the following two conditions: $10\underline{e} - w \geq 0$ and $w - c(\underline{e}) \geq 0$. 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). All major results hold when we focus only on the IR/IC contracts. We analyze in detail the IR/IC contracts in Appendix B and the non-IR/IC contracts in Appendix C.

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.

4.1. 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 over periods 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 (0.14 versus -0.50; p-value = 0.33, $n_1 = 9$, $n_2 = 9$), Effort-Only versus Effort-Shock (0.14 versus -0.18; p-value = 0.82, $n_1 = 9$, $n_2 = 9$), and Outcome-Only versus Effort-Shock (-0.50 versus -0.18; p-value = 0.60, $n_1 = 9$, $n_2 = 9$).[21]

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. The model suggests that the conditional adjustment (that is, conditional on wage, effort, desired effort, and possibly the shock) may differ across treatments, and it can also give some insight into the motivations of the subjects. Specifically, if principals reciprocate based on effort, the adjustment should not vary with the shock in the Effort-Shock treatment (Prediction 1), whereas the adjustment should vary with the shock if principals reciprocate based on outcome (Prediction 2). More generally, 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 first test whether the effort (minus desired effort) varies across shock level in the Effort-Shock treatment. Table 2 reports the average adjustment in the Effort-Only and Effort-Shock treatments, as a function of whether *effort – desired effort* is negative, zero, or positive and for the Effort-Shock

[21] We have also checked the robustness of these results using panel regression analysis. Specifically, we have estimated 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 are an inverse of *period*, *wage*, *effort – desired effort* (in Effort-Only and Effort-Shock) and *outcome – desired effort* (in Outcome-Only), as well as treatment dummies. All regressions indicate no significant difference in adjustment level between the three treatments. The estimation results are available in Appendix D.

treatment whether the *shock* is negative, zero, or positive. The results provide a preliminary basis for rejecting the effort-based reciprocity hypothesis, as adjustments appear to vary as the shock varies. Focusing on the case where *effort – desired effort* is negative (since N is high enough in this case to support statistics), the average adjustment made after a negative shock is -1.11, whereas the average adjustment made after a positive shock is -0.32. This difference is statistically significant (Mann-Whitney, p-value = 0.02). In fact, even the average adjustment made after a zero shock, -0.55 is marginally greater than the average adjustment made after a negative shock (p-value = 0.10).[22] Since the average adjustment varies as the shock varies, we can reject the hypothesis that principals reciprocate based solely on effort.

Next, we test whether principals condition their adjustments based on previous actions. Table 3 reports the estimation results of different panel models where individual subjects represent the random effects, and the standard errors are calculated using a bootstrap method.[23] 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).[24] 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.[25]

While these results suggest that principals 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 0.38 for every unit of effort given (relative to desired effort), whereas the marginal increase is only 0.23 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 reaffirms the conjecture 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. Consistent with our previous findings in Table 2, we find that the *adjustment* and the *shock* variables are positively correlated, suggesting that principals punish or reward agents based in part on *outcomes*. In fact, a

[22] The differences in average adjustment across different shock levels are not significant when *effort – desired effort* is 0 or positive (although in two cases the p-value comes very close to significance), but this is likely due to the low number of observations in these cases.

[23] Since we have a relatively low number of clusters, we have used bootstrap method to calculate the standard errors within each cluster (Cameron et al., 2008).

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

[25] Table 2 suggests that principals pay a significant, and possibly discontinuous, premium for having contracts fulfilled (i.e., *effort* ≥ *desired effort*). We test this by re-analyzing the results in Table 3, replacing the "*effort*/*outcome – desired effort*" variables with dummies for whether the contract was fulfilled. The coefficients on these dummies are highly significant in all 5 specifications and their magnitudes range from 1.61 to 2.06. The coefficient on the shock variable in column (4) remains highly significant, and none of the results reported above are qualitatively altered.

comparison of specifications (3) and (5) suggests that principals respond similarly (0.21 versus 0.23) to an increase in *outcome* regardless of whether or not the effort is observed.[26]

Result 1: There is no significant difference in the unconditional *adjustment* level between treatments. The adjustment level varies positively with effort and the shock.

4.2. Effort

We next consider the effort that the agent chooses in stage 2. Figure 3 displays the average effort over periods by treatment, while Figure 4 displays the distribution of effort by treatment. Based on the Wilcoxon rank-sum test, we find that the average effort in the Effort-Only treatment is higher than in the Outcome-Only treatment (6.40 versus 4.69; p-value = 0.02, $n_1 = 9$, $n_2 = 9$) and the Effort-Shock treatment (6.40 versus 4.69; p-value = 0.01, $n_1 = 9$, $n_2 = 9$). On the other hand, the average effort is not different between the Outcome-Only and Effort-Shock treatments (4.69 versus 4.69; p-value = 0.82, $n_1 = 9$, $n_2 = 9$).[27] These results are consistent with Predictions 2 and 3 of the model, which suggest that the effort given in the Effort-Only treatments is higher than in the other two treatments if principals reciprocate based on the outcome of the game play.

The intuition laid out previously indicates 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 4 reports the estimation results of different panel models, where the dependent variable in all specifications

[26] 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 papers by Abbink et al. (2000), Fehr and Gächter (2000) and Baumeister et al. (2001), many studies have shown that "negative" reciprocity is stronger than "positive" reciprocity (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., $(\textit{effort} - \textit{desired effort})^2$ and $(\textit{outcome} - \textit{desired effort})^2$ are negative) and negative reciprocity decreases at a diminishing rate (i.e., $(\textit{effort} - \textit{desired effort})^2$ and $(\textit{outcome} - \textit{desired effort})^2$ are positive). These results are not always statistically significant, however.

[27] We have also checked the robustness of these results using a panel regression analysis controlling for individual subject effects, re-matching groups, learning, *wage*, and *desired effort*. The regression results corroborate our main findings: effort is greater in Effort-Only than in the other two treatments, but there is no difference between Effort-Shock and Outcome-Only treatments. The estimation results are available in Appendix D.

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*, 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 4 indicate that the magnitude of the effect of *desired effort* on the *effort* chosen differs between treatments. In the Effort-Only treatment, principals receive 11% of each additional unit of *effort* they desire (and this is statistically significant), whereas the magnitude is 8% in the Effort-Shock treatment and 5% in the Outcome-Only treatment (although neither are statistically significant). These results suggest that agents form reasonably correct beliefs regarding how principals will act in the adjustment period. Table 3 suggests 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 (relative to the Effort-Only treatment) in order to avoid any chance of being perceived as underperforming the desired effort. Whether agents choose effort greater than desired effort more frequently in the Effort-Shock treatment relative to the Effort-Only treatment depends on the degree to which agents believe that principals reward/punish based on *outcome* relative to *effort*. If agents believe that principals reciprocate based solely on effort relative to desired effort (with a discontinuity at equality), they should not give effort above desired effort more frequently in the Effort-Shock treatment than in the Effort-Only treatment. They should give such extra effort in the Effort-Shock treatment, however, if outcomes are the primary driver of reciprocity, since negative shocks are possible. 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 (0.06 versus 0.17; p-value = 0.01, $n_1 = 9$, $n_2 = 9$), it is marginally lower in the Effort-Only than in the Effort-Shock treatment (0.06 versus 0.13; p-value = 0.10, $n_1 = 9$, $n_2 = 9$), and not significantly different between the Outcome-Only and Effort-Shock treatments (0.17 versus 0.13; p-value = 0.53, $n_1 = 9$, $n_2 = 9$). This further corroborates the finding that agents expect principals to exhibit outcome-based reciprocity.

Moreover, 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) more often in the Effort-Only treatment than in the other two treatments. There are no shocks in this treatment, so effort is equal to outcome, whereas shocks distort the mapping from effort to outcome in the other two treatments. 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 (0.29 versus 0.10; p-value < 0.01, $n_1 = 9$, $n_2 = 9$) and the Effort-Shock treatment (0.29 versus 0.17; p-value = 0.03, $n_1 = 9$, $n_2 = 9$). There is no statistically significant difference between the Outcome-Only and Effort-Shock treatments (0.10 versus 0.17; p-value = 0.15, $n_1 = 9$, $n_2 = 9$).

Result 3: Effort levels respond positively to *desired effort* in the Effort-Only and Effort-Shock treatments. There is a greater probability that effort exceeds *desired effort* in the Outcome-Only and Effort-Shock treatments 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.

4.3. 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.

Figures 6 and 8 display the average wage and desired effort over periods and by treatment, while Figures 7 and 9 display 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 (41.14 versus 33.45; p-value = 0.05, $n_1 = 9$, $n_2 = 9$) and Outcome-Only treatment (41.14 versus 33.85; p-value = 0.08, $n_1 = 9$, $n_2 = 9$). Similarly, we find that the average desired effort is higher in the Effort-Only treatment than in the Effort-Shock treatment (8.95 versus 7.63; p-value = 0.02, $n_1 = 9$, $n_2 = 9$) and the Outcome-Only treatment (8.95 versus 7.63; 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 = 0.57 and 0.89, respectively).[28]

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 statistically 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 in part 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 4, which indicated that effort responds significantly to desired effort in the Effort-Only treatment and not in the other two treatments. 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, $\rho = 0.54$, indicating that higher wages are associated with higher desired effort.

[28] We have also checked the robustness of these results using panel regression analysis, controlling for individual subject effects, re-matching groups, and learning. The regression results corroborate our main findings: wage and desired effort are greater in Effort-Only than in the other two treatments, but there is no difference between Effort-Shock and Outcome-Only treatments. The estimation results are available in Appendix D.

4.4. Payoffs and Welfare

As a consequence of higher wage and higher effort, the Effort-Only treatment generates a significantly higher payoff to the principal than the Outcome-Only treatment (20.91 versus 11.23; p-value = 0.05, $n_1 = 9$, $n_2 = 9$) and the Effort-Shock treatment (20.91 versus 12.04; 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 (11.23 versus 12.04; p-value = 0.89, $n_1 = 9$, $n_2 = 9$). When comparing payoffs of agents, we find no significant differences between the three treatments (all p-values > 0.48).

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 (36.62 versus 29.67; p-value = 0.05, $n_1 = 9$, $n_2 = 9$) and Effort-Only versus Effort-Shock (36.62 versus 27.64; 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 (27.64 versus 29.67; p-value = 0.31, $n_1 = 9$, $n_2 = 9$).

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

5. 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 in part on the basis of the *outcome* of the exchange, even if part of the outcome is determined by forces outside the control of the other party. 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).

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.[29] 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 our results are not being driven by agents "hiding behind randomness", where they give less effort when they can blame a bad outcome on a negative shock (even if the shock ended up not being negative). 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.[30] But this motivation cannot account for the multitude of differences we see between the Effort-Only and Effort-Shock treatments; since agents cannot "hide behind randomness" in the latter treatment, they should not act differently than when there is no randomness. Indeed, any explanation that cannot account for differences based on the observability of actions cannot explain our treatment differences.

[29] 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.

[30] 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 Effort-Only and Effort-Shock treatments.

What, then, can explain our results? First, our results are consistent with principals exhibiting outcome-based reciprocity. However, this simply means that we cannot reject this motivation as driving actions; it is possible that other motivations are at work as well. To this end, we believe that there are two other, 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 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 uncertainty. 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 shock 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 probability of being perceived as choosing 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 partially mitigated by the presence of shocks, whether or not the shocks are observed.

References

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner. (2000). "The moonlighting game: An experimental study on reciprocity and retribution." *Journal of Economic Behavior and Organization* 42(2): 265-277.

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(4): 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., and Vohs, K. (2001). "Bad is stronger than good." *Review of General Psychology* 5(4): 323-370.

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

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. (2008). "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics* 90(3): 414-427.

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(2): 203-231.

Charness, Gary, and Uri Gneezy. (2008). "What's in a name? Anonymity and social distance in dictator and ultimatum games." *Journal of Economic Behavior and Organization* 68(1): 29-35.

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*, pp. 229-330.

Charness, Gary and Martin Dufwenberg. (2006). “Promises and partnership.” *Econometrica* 74(6): 1579-1601.

Charness, Gary, and D. I. Levine. (2007). “Intention and Stochastic Outcomes: An Experimental study.” *Economic Journal* 117(522): 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(2): 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(2): 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(2): 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.

Gneezy, Uri, and John A. List. (2006). “Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments.” *Econometrica*, 74(5): 1365-1384.

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

Gurdal, M., Miller, J., and Rustichini, A. (2013). “Why Blame?” Working Paper.” *Journal of Political Economy* 121(6): 1205-1247.

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(2): 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, and 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(2): 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.

List, John A. (2007). “On the interpretation of giving in dictator games.” *Journal of Political Economy* 115(3): 482-493.

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(3): 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(2-3): 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(8): 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. (2015). “Effort and Luck: The Effects of Information and Culture on Giving” 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(5): 751-765.

Xiao, Erte and Howard Kunreuther. (2015). "Punishment and Cooperation in Stochastic Prisoner’s Dilemma Game." *Journal of Conflict Resolution*, forthcoming.

Figure 1: Average Adjustment by Period.

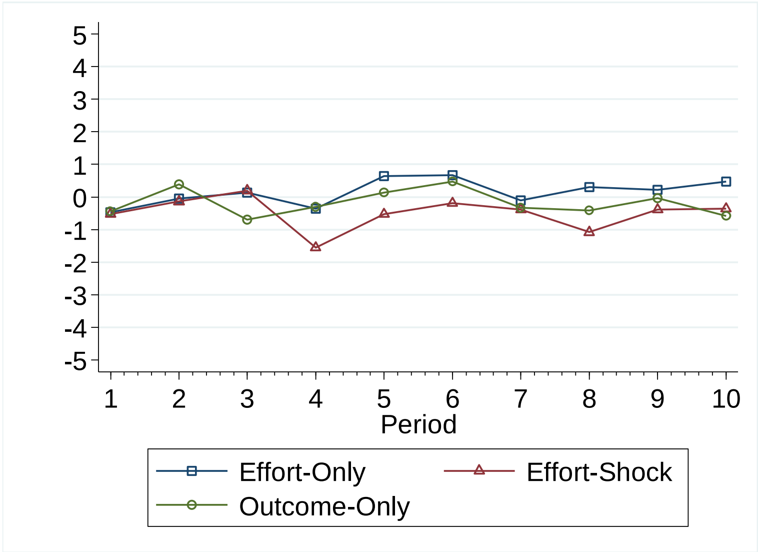


Figure 2: Distribution of Adjustment.

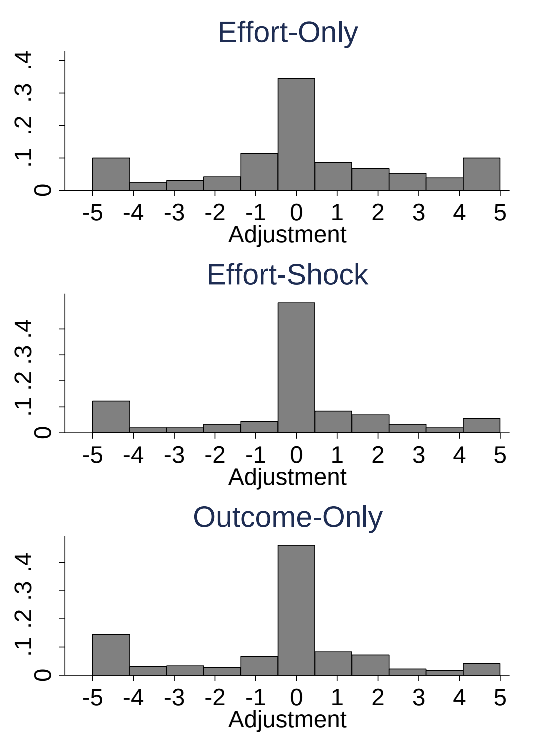


Figure 3: Average Effort by Period.

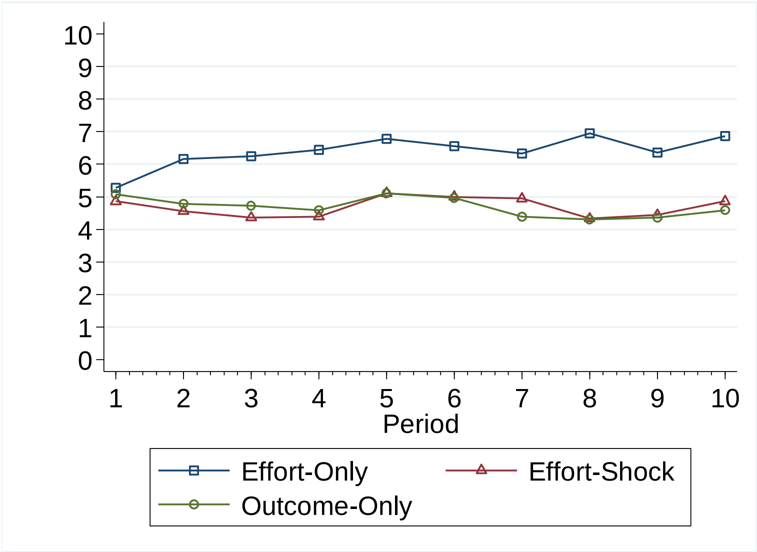


Figure 4: Distribution of Effort.

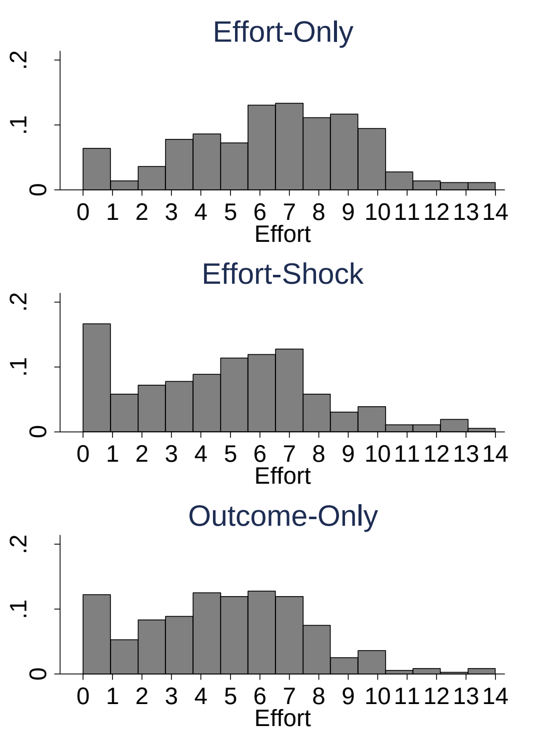


Figure 5: Distribution of Effort – Desired Effort.

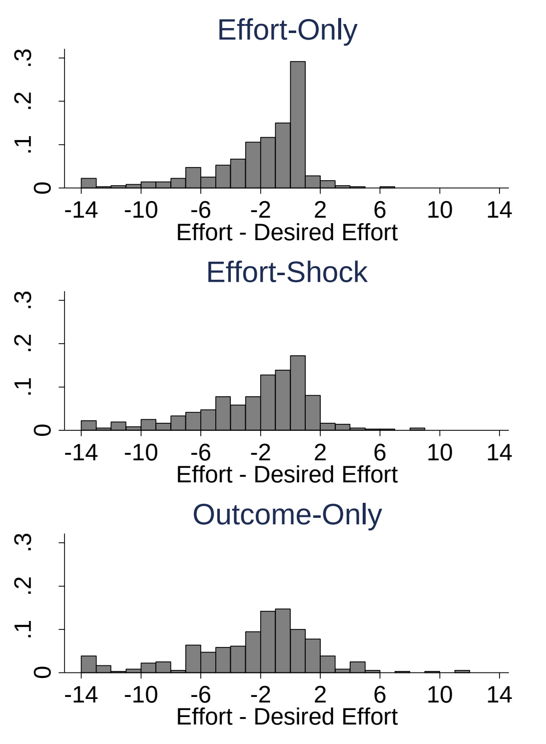


Figure 6: Average Wage by Period.

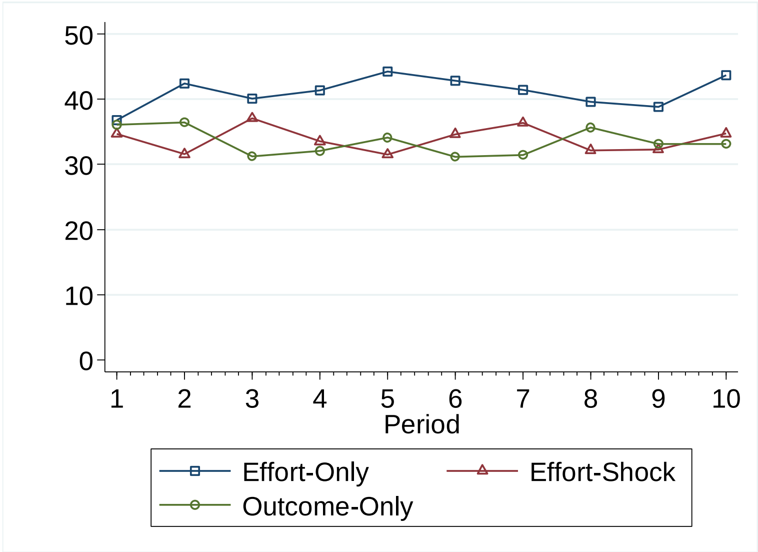


Figure 7: Distribution of Wage.

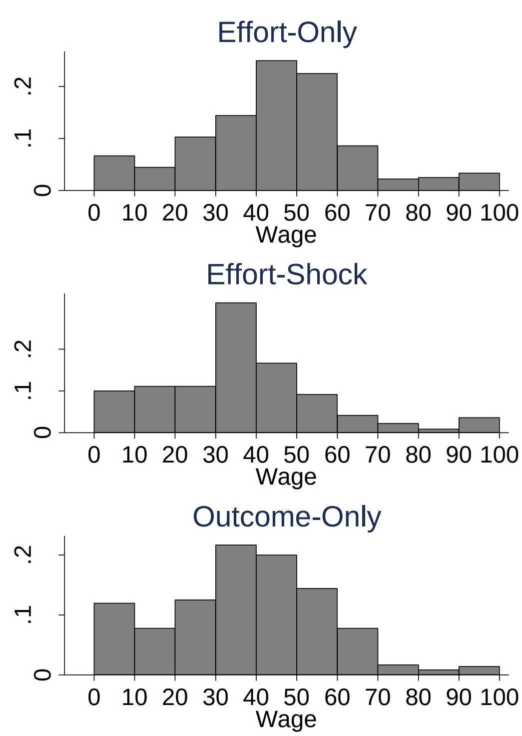


Figure 8: Average Desired Effort by Period.

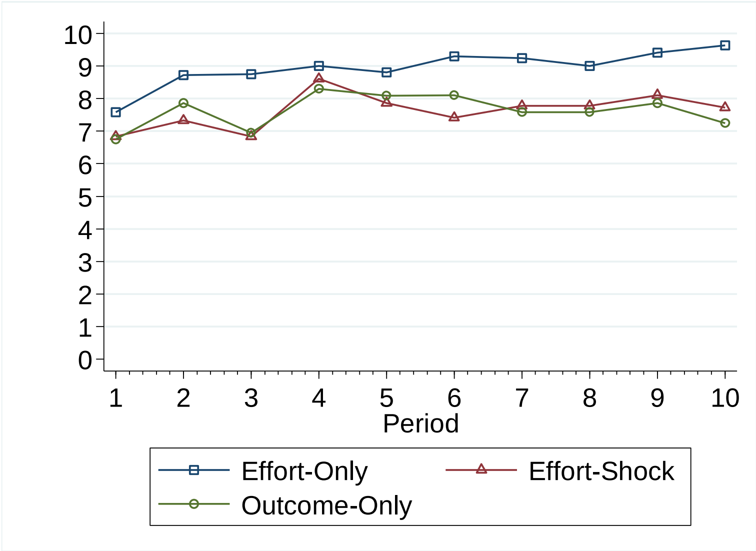


Figure 9: Distribution of Desired Effort.

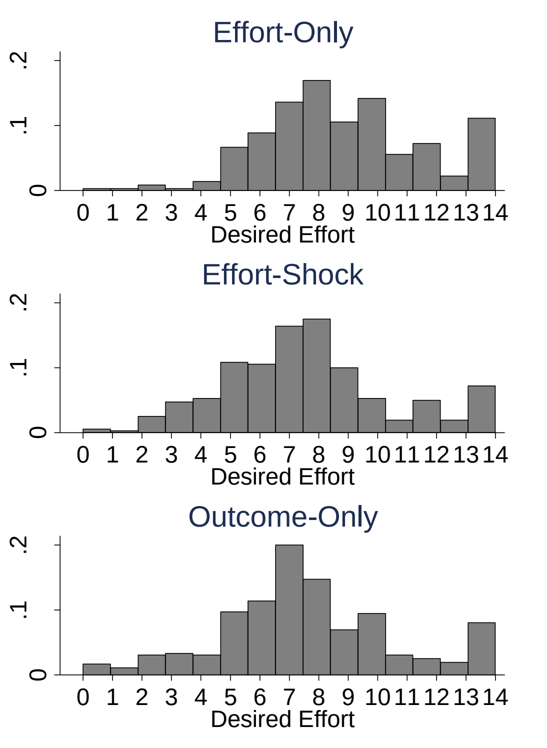


Table 1: Summary Statistics.

Treatment	Wage	Desired Effort	Effort	Outcome	Adjustment	Principal's Payoff	Agent's Payoff	Total Welfare
Effort-Only	41.14	8.95	6.40	6.40	0.14	20.91	15.71	36.62
	(3.22)	(0.31)	(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	16.41	27.64
	(2.98)	(0.34)	(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 are based on 9 independent observations.

Table 2: Average Adjustment given Effort minus Desired Effort and Shock.

	effort – desired effort		
	Negative	*Zero*	*Positive*
Effort-Only	-0.57	1.55	1.10
N	235	105	20
Effort-Shock	-0.70	0.90	1.22
N	252	62	46
Negative Shock	-1.11	1.00	0.71
N	109	23	17
Zero Shock	-0.55	-0.13	1.50
N	44	16	14
Positive Shock	-0.32	1.52	1.53
N	99	23	15

Table 3: 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**	-0.01	-0.01**	-0.01**
[wage]	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
effort – desired effort	0.38***	0.23***		0.23***	
[effort gap]	(0.08)	(0.05)		(0.05)	
outcome – desired effort			0.21***		0.23***
[outcome gap]			(0.06)		(0.05)
shock				0.23***	
[random number]				(0.09)	
period	-1.12***	-0.45	-0.21	-0.51	-0.51
[inverse period]	(0.49)	(0.37)	(0.58)	(0.39)	(0.39)
constant	2.01***	0.96***	0.38	1.01***	1.00***
[constant term]	(0.82)	(0.33)	(0.58)	(0.33)	(0.33)
N	360	360	360	360	360
Clusters	9	9	9	9	9
Overall R-squared	0.18	0.1	0.08	0.12	0.12

* significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors in parenthesis are clustered at the group level and are calculated using a bootstrap method.

Table 4: Panel Models of Effort.

Specification	(1)	(2)	(3)
Treatments	*Effort-Only*	*Effort-Shock*	*Outcome-Only*
Dependent variable	*Effort*	*effort*	*effort*
wage	0.09***	0.07***	0.06***
[wage]	(0.01)	(0.01)	(0.01)
desired effort	0.11***	0.08	0.05
[desired effort]	(0.05)	(0.06)	(0.05)
period	-0.91**	0.39	0.09
[inverse period]	(0.43)	(0.64)	(0.50)
constant	1.93***	1.52***	2.21***
[constant term]	(0.60)	(0.44)	(0.42)
N	360	360	360
Clusters	9	9	9
Overall R-squared	0.43	0.25	0.18

* significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors in parenthesis are clustered at the group level and are calculated using a bootstrap method.

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.

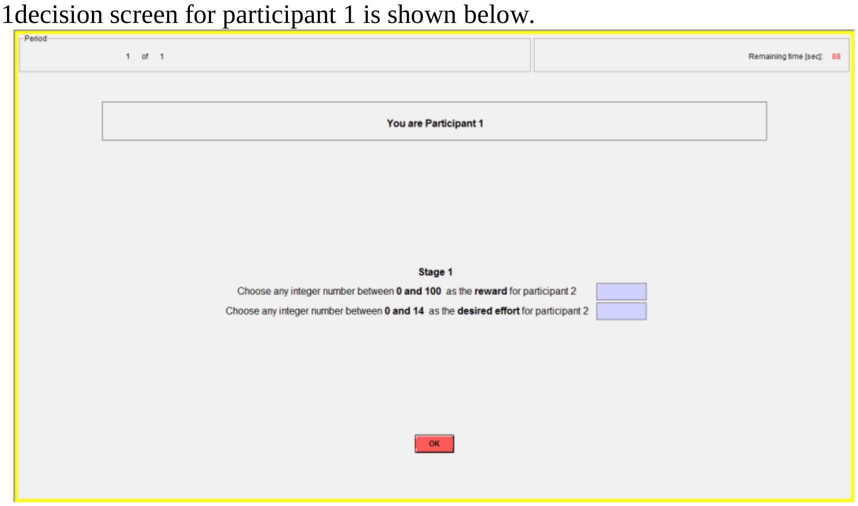
It is very important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid.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

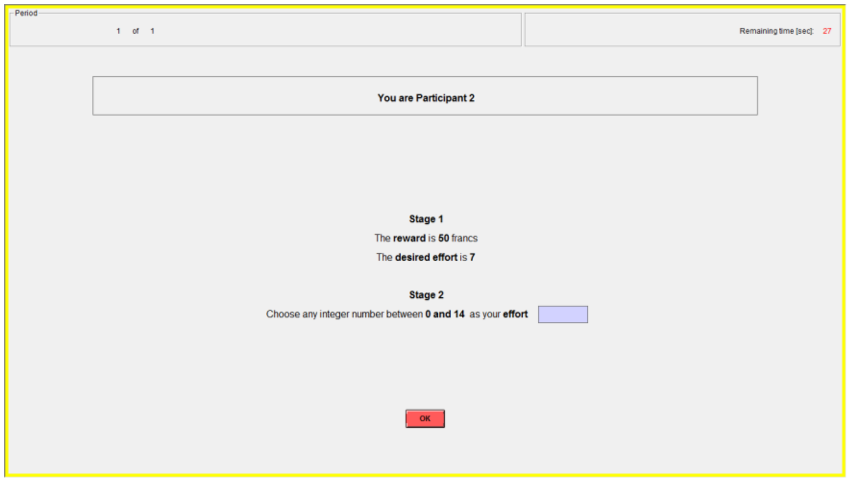
Each period will proceed in three stages. In Stage 1, **participant 1** will choose a **reward** (any integer number between **0 and 100**) and a **desired effort** (any integer number between **0 and 14**) for participant 2. An example of Stage 1decision screen for participant 1 is shown below.



[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:

$$\text{participant 2's performance} = \text{effort} + \text{random number}$$

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

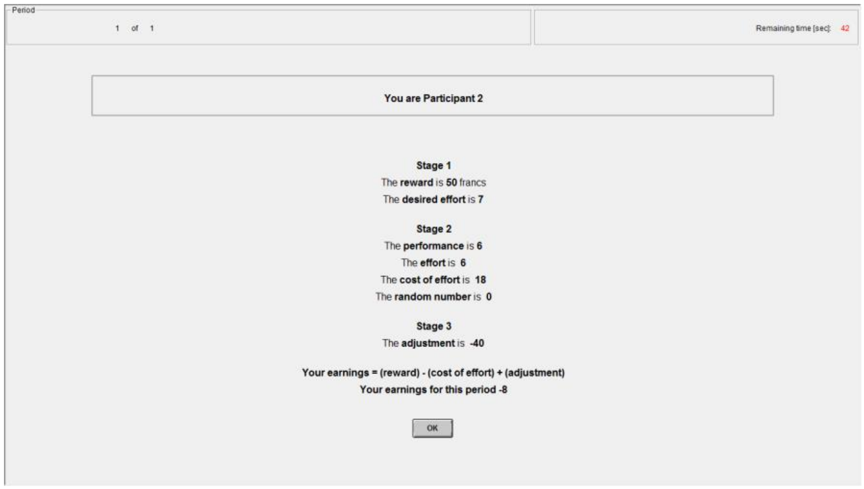
Assume the following scenario. In the first stage, participant 1 chooses a reward of 50 and a desired effort of 7. In the second stage, participant 2 chooses an effort of 6. Then the computer randomly selects 2 as a random number, so the performance of participant 2 is 8 (6+2). Then the computer displays 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): IR/IC Contracts

The analysis in Section 4 is based on all contracts. Here, we focus on the IR/IC contracts. The majority of contracts observed in our experiment satisfy IR/IC (62%). 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). Table B1 provides the summary statistics across all three treatments.

When examining adjustments in stage 3, we find no significant difference 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 = 0.20, $n_1 = 9$, $n_2 = 9$), and Outcome-Only versus Effort-Shock (-0.62 versus -0.51; p-value = 0.96, $n_1 = 9$, $n_2 = 9$).

When examining effort in stage 2, 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 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, $n_1 = 9$, $n_2 = 9$). When examining the probability of effort exceeding the desired effort, we find that it 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, $n_1 = 9$, $n_2 = 9$). When examining the probability of agents choosing efforts exactly specified by the contract, we find that it is significantly higher in the Effort-Only treatment 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.

When examining wage and desired effort in stage 1, we find that they are the highest in the Effort-Only treatment. 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, 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). 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).

Table B1: Summary Statistics: IR/IC Contracts.

Treatment	Wage	Desired Effort	Effort	Outcome	Adjustment	Principal's Payoff	Agent's Payoff	Total Welfare
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)

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

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

The analysis in Section 4 is based on all contracts. Here, we focus on the non-IR/IC contracts. 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 individually rational. Table C1 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 C1), 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 C1), 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 C1 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 C1: Summary 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.

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

Table D1 shows 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 are an inverse of *period*, *wage*, *effort – desired effort* (in *Effort-Only* and *Effort-Shock*) and *outcome – desired effort* (in *Outcome-Only*), as well as treatment dummies. All regressions indicate no significant difference in adjustment level between the three treatments (see the treatment dummies).

Table D1: Panel Models of Adjustments.

Specification	(1)	(2)	(3)
Treatments	*Effort-Only & Effort-Shock*	*Effort-Only & Outcome-Only*	*Effort-Shock & Outcome-Only*
Dependent variable	*adjustment*	*adjustment*	*adjustment*
Effort-Only	0.28	0.62	
[Effort-Only treatment dummy]	(0.42)	(0.41)	
Outcome-Only			-0.34
[Outcome-Only treatment dummy]			(0.25)
wage	-0.01**	-0.01	-0.01**
[wage]	(0.00)	(0.01)	(0.00)
effort – desired effort	0.30***		
[effort gap]	(0.05)		
outcome – desired effort		0.28***	0.22***
[outcome gap]		(0.05)	(0.04)
period	-0.81***	-0.69*	-0.36
[inverse period]	(0.30)	(0.41)	(0.32)
constant	1.24***	0.80*	0.87***
[constant term]	(0.37)	(0.44)	(0.32)
N	720	720	720
Clusters	18	18	18

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

Table D2 shows 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 *effort* and the independent variables are an inverse of *period*, *wage*, and *desired effort*, as well as treatment dummies. The regression results corroborate our main findings (see the treatment dummies): effort is greater in Effort-Only than in the other two treatments, but there is no difference between Effort-Shock and Outcome-Only treatments.

Table D2: Panel Models of Effort.

Specification	(1)	(2)	(3)
Treatments	*Effort-Only & Effort-Shock*	*Effort-Only & Outcome-Only*	*Effort-Shock & Outcome-Only*
Dependent variable	*effort*	*effort*	*effort*
Effort-Only	0.95**	1.03**	
[Effort-Only treatment dummy]	(0.41)	(0.49)	
Outcome-Only			-0.03
[Outcome-Only treatment dummy]			(0.39)
wage	0.08***	0.08***	0.07***
[wage]	(0.01)	(0.01)	(0.01)
desired effort	0.09**	0.09**	0.07
[desired effort]	(0.04)	(0.05)	(0.04)
period	-0.30	-0.43	0.26
[inverse period]	(0.40)	(0.32)	(0.42)
constant	1.30***	1.49***	1.82***
[constant term]	(0.37)	(0.45)	(0.42)
N	720	720	720
Clusters	18	18	18

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

Table D3 shows 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 *wage* and the independent variables are an inverse of *period*, and treatment dummies. The regression results corroborate our main findings (see the treatment dummies): wage is greater in Effort-Only than in the other two treatments, but there is no difference between Effort-Shock and Outcome-Only treatments.

Table D3: Panel Models of Wage.

Specification	(1)	(2)	(3)
Treatments	*Effort-Only & Effort-Shock*	*Effort-Only & Outcome-Only*	*Effort-Shock & Outcome-Only*
Dependent variable	*wage*	*wage*	*wage*
Effort-Only	7.69*	7.29*	
[Effort-Only treatment dummy]	(4.26)	(3.83)	
Outcome-Only			0.40
[Outcome-Only treatment dummy]			(3.65)
period	-0.40	-2.04	2.24
[inverse period]	(3.12)	(3.09)	(3.45)
constant	33.57***	34.45***	32.80***
[constant term]	(3.43)	(2.92)	(3.46)
N	720	720	720
Clusters	18	18	18

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

Table D4 shows 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 *desired effort* and the independent variables are an inverse of *period*, and treatment dummies. The regression results corroborate our main findings (see the treatment dummies): wage is greater in Effort-Only than in the other two treatments, but there is no difference between Effort-Shock and Outcome-Only treatments.

Table D4: Panel Models of Desired Effort.

Specification	(1)	(2)	(3)
Treatments	*Effort-Only & Effort-Shock*	*Effort-Only & Outcome-Only*	*Effort-Shock & Outcome-Only*
Dependent variable	*desired effort*	*desired effort*	*desired effort*
Effort-Only	1.31***	1.32***	
[Effort-Only treatment dummy]	(0.44)	(0.38)	
Outcome-Only			-0.01
[Outcome-Only treatment dummy]			(0.41)
period	-1.46***	-1.57**	-1.11*
[inverse period]	(0.54)	(0.70)	(0.60)
constant	8.06***	8.09***	7.96***
[constant term]	(0.41)	(0.38)	(0.42)
N	720	720	720
Clusters	18	18	18

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