Inter-Firm Social Dilemmas with Agency Risk*

Timothy N. Cason
Purdue University
cason@purdue.edu

Lana Friesen
University of Queensland
l.friesen@uq.edu.au

Lata Gangadharan
Monash University
Lata.Gangadharan@monash.edu

August 2020

Abstract: Many social dilemmas involve decisions made by firms. We design a laboratory experiment that represents firms' principal-agent problem and includes an inter-firm social dilemma and stochastic agent performance. Agents' unobservable effort affects the likelihood of a bad outcome occurring, such as a regulatory violation. This harms the agent's principal but can also damage others, thus creating an inter-firm social dilemma. In our baseline treatment, we omit the agency problem, and principals make their "firm's" effort decision directly. In the second treatment, principals can only offer an unconditional wage contract to their agent, although a non-contractual (ex-post) bonus can be paid. In a third treatment, principals can condition wages on the stochastic outcome, and a fourth treatment combines the conditional wage with a non-contractual bonus. We find that principals use a combination of a conditional wage and the non-contractual (ex-post) bonus to help overcome the agency problem and incentivize agents to choose higher effort. Fixed wage, unconditional contracts lead to significantly lower effort levels, even when augmented with bonuses. Similarly, conditional contracts on their own also perform poorly. Only the combination of conditional wage contracts and discretionary bonuses is effective in limiting agency risk to address the inter-firm social dilemma problem.

Keywords: Experiments; Gift Exchange; Principal-Agent; Externalities; Reciprocity

JEL Classification: C91; D90; H23; Q50

* We thank Daniel Woods for excellent research assistance and the Krannert School of Management for funding the subject payments. We are also grateful for helpful comments provided by three anonymous referees, as well as Glenn Harrison, John List, Jean Paul Rabanal, Tim Shields, Roman Sheremeta, and conference and seminar audiences at the Australia New Zealand Workshop in Experimental Economics, Economic Science Association meetings, Southern Economic Association, World Congress for Environmental and Resource Economics, Behavioral and Experimental Public Choice Workshop, Workshop on Resource Security and Economic Sciences, Xiamen, Bath, Griffiths and Monash Universities, Virginia Tech and the University of Mannheim.

1. Introduction

Social dilemmas are defined by a conflict between private (self-) interest and potential gains to society from mutual cooperation. Important examples include public good provision through voluntary contributions and common pool resource management. To date most research on social dilemmas has focused on decisions made by individuals. Yet many important social dilemmas involve decisions made by firms who face challenges with agency risk. Concurrently a large literature considers principal-agent problems within the firm and demonstrates the challenges of aligning incentives between owners, managers and workers. This paper brings together the two influential research agendas of social dilemmas and principal agent problems by studying whether agency risk exacerbates social dilemma problems, and which contract forms, such as state-contingent wages or nonbinding bonuses, might improve decision-making and social outcomes.

Consider a situation in which firms have to comply with regulations. Within each firm, an agency problem can arise in many different ways, such as between owners and managers, managers and workers, or firms and their subcontractors. Regardless, if the principal wants to comply with relevant regulations, she must rely on the agent to implement the necessary actions. Usually if (civil) penalties are imposed, they affect the firm (principal) rather than the agent and/or the individual is indemnified from or insured against personal penalties, thus creating a classic problem of moral hazard. The fallout from the global financial crisis exemplifies this issue: while banks and other financial institutions paid billions to settle civil cases, few of the individuals responsible were held personally liable.[1]

The key novelty of our research is to embed the firm's agency problem in a new variant of a social dilemma, which we refer to as an inter-firm social dilemma: if the bad outcome (e.g. a regulatory violation) occurs, then not only do the principal's earnings fall but other principals in the group also suffer damages. This externality could arise through direct harm to others or due to increased costs from greater regulatory scrutiny applied to the entire industry. The financial crisis, for example, led to several new stringent measures for all participants in the financial industry (Claessens and Kodres, 2014).

To examine how the intra-firm principal-agent relationship influences the inter-firm social dilemma and whether contracts can be tailored to reduce agency risk, we design a laboratory

[1] This is consistent with the legal doctrine *respondeat superior*. In the United States, for example, employers are often liable for acts of employees performed over the course of their employment.

experiment that compares different kinds of agency contracts. Our experiment utilizes a novel variation of the well-established gift exchange framework (e.g., Fehr et al, 1993). Although in the classical principal-agent model the owner cannot perfectly observe the employee's effort, in most previous gift exchange experiments both parties can observe effort even when it is non-contractible. In contrast, in our experiment, agents' effort choices determine the likelihood of a bad outcome occurring (such as a detected regulatory violation). Besides being non-contractible, effort is therefore also unobservable to the principal—consistent with the prototypical principal-agent model so influential in management and economics.

Our goal is to overlay the agency risk on a social dilemma problem. If other firms can impose negative externalities on the principal, then paying a high wage to her agent when he avoids an accident can still expose her to substantial risk. This puts an upper limit on the wage a principal can pay to reward the agent for a good outcome, and this limit may be insufficient to induce the desired behavior. Thus, the inter-firm social dilemma could constrain the effectiveness of intra-firm incentive contracts. The ability to distribute bonuses ex post can help solve this problem. While bonuses are non-binding, since they are flexible they give the principal discretionary power over how to distribute the final gains after stochastic outcomes and externalities are realized. In our study we keep the inter-firm social dilemma constant across treatments, while varying the available payment forms to compare the effectiveness of different kinds of intra-firm incentive contracts.

We consider four treatments in a between-subjects experimental design to explore the effectiveness of different types of incentive payments, all of which are set in an inter-firm social dilemma. The baseline treatment eliminates the agency problem because principals make their own effort decision. This gives us a measure of principals' preferences regarding their concern for other firms in the inter-firm social dilemma. We contrast this with three agency treatments that explore alternative contractual formats for incentivizing agent performance. Agency contracts designed to mitigate the impact of the social dilemma could take different forms. Some set out explicit conditions to create contractual incentives for the agent to take the preferred action, as seen in a wide variety of incentive pay plans. Non-binding incentives such as bonuses, on the other hand, provide the principal with greater flexibility in rewarding the agent for the preferred action, such as in "spot awards" used by over one-third of firms (Milkovich et al., 2001, Ch. 10). Our experimental design allows us to explore and compare the effectiveness of these different contractual measures.

In our second treatment, principals can only offer an unconditional wage contract to their agents, although they can pay a non-contractual (ex-post) bonus. This contract is closest in spirit to the traditional gift exchange framework. A third treatment examines whether an explicit contract in which wages can be conditioned on the outcome (but not effort, which is unobservable) activates sufficient gift exchange motivations to encourage the agent to undertake the principal's preferred effort. Our fourth treatment combines the different contractual features and allows the principal to offer both conditional wages (contracted on the outcome) and bonuses (non-contractual). This treatment can help us understand if principals use the option to give a bonus in situations where wages can be explicitly conditioned on the outcome and the tradeoffs they face in the presence of the social dilemma.

We find that principals use a combination of a conditional wage and the non-contractual (ex-post) bonus to help overcome the agency problem and incentivize agents to choose higher efforts. By contrast, unconditional contracts (that depend exclusively on gift exchange relationships) and conditional contracts that do not allow for bonuses to be offered lead to significantly lower effort levels. Only when both conditional wage contracts and ex-post bonuses are possible does agency risk not substantially exacerbate the social dilemma problem. Our finding that fixed wage contracts lead to worse outcomes, even with (non-contractual) bonuses, indicates that bonuses alone are ineffective for mitigating agency problems in this social dilemma environment with unobservable effort. Similarly, explicit contractual wages contingent on the outcome, but with no allowance for bonuses, also lead to worsening of the social dilemma.

Our findings help reconcile the use of two theoretically distinct contractual features, conditional wages and bonuses. Ex-post bonuses appear to be used as a risk management strategy by principals, for both themselves and the agent. By paying a lower conditional wage, principals can reduce their own risk because in this social dilemma high group damage can arise from others' outcomes. The bonus allows them to reduce the risk they face by paying the agent a higher bonus in case of a good outcome. In the absence of bonuses, the principal must offer a wide spread in the conditional wages to incentivize the agent to take the appropriate effort choice, but this makes the agent bear all the risk. The flexibility to combine binding conditional wages with discretionary bonuses helps principals manage risk and at least partially overcome the agency problem.

The central innovation of our experiment is embedding the agency problem within an inter-firm social dilemma, with bad outcomes affecting not only the agent's principal but also others. Researchers have examined many dimensions of individual social dilemmas, and a few papers (such as Charness et al., 2007, and Abbink et al., 2010) have studied group decisions in social dilemmas. To our knowledge, however, no other studies consider inter-firm social dilemmas in a principal-agent gift exchange setting. Two recent experiments examine situations where agent choices affect others outside of the direct agency relationship. Dijk and Holmen (2017) use a gift exchange experiment where agent choices can *help* others outside of the agency relationship. They show that both effort levels and efficiency are higher when the principals' earnings are donated to a charity compared to the usual case where they are retained by the principal, but this is mitigated when the principals' earnings are paid to another experimental subject in the session. These results illustrate how intra-firm cooperation might be sustained with positive externalities, but they do not consider the case of inter-firm interactions. d'Adda et al. (2017) consider inter-firm interactions but in a very different context where worker dishonesty helps their own firm's profits but reduces the profits of other firms. They focus on the influence of leaders on worker dishonesty and show that leaders influence behavior using a combination of bonuses and statements. In contrast, in our setting workers' shirking harms both their own firm and other firms in their group. The main novelties of our study are to consider whether agency risk exacerbates inter-firm social dilemmas, and to study both explicit and non-binding contracts and their combination as potential solutions. These two aspects remain largely unexplored in the literature.

Whether a reciprocity-based relationship, such as a gift exchange contract, can improve effort in this important and relevant environment of an inter-firm social dilemma is an open question. This experimental design enables us to examine the impact of different types of reciprocity. First, the inter-firm social dilemma might elicit reciprocity between principals, which our baseline treatment will reveal. Second, the agency treatments may trigger reciprocal motivations between the principal and agent as in the classic gift exchange environment. However, the agency problem might weaken the degree to which social preferences help move choices towards the social optimum in the social dilemma since the principal has incomplete control of the agent's choice thus providing the principal with some "moral wiggle room." Erat (2013), for example, shows that many people prefer to avoid lying themselves but rather delegate the choice to an agent, especially when lying is particularly harmful to others.[2]

Our finding that principals do not fully utilize the conditional contract but instead use a combination of a conditional wage and the ex-post bonus relates to findings in the literature that effort is highest when principals forgo the opportunity to use a penalty or incentive contract and instead rely on trust. Fehr and Rockenbach (2003) and Fehr and List (2004) both find that trustworthiness increases when principals choose a trust contract when they could have chosen a penalty contract. Fehr and List (2004) find this among both students and CEOs. Similarly, Fehr et al. (2007) find that effort significantly increases in a gift exchange experiment when principals *choose* a bonus contract rather than an incentive contract.[3] Both papers attribute this result to the intention conveyed by the choice, with trust sustaining reciprocity while explicit penalties convey hostile intentions that undermine reciprocity.[4] In our experiment, principals do not rely exclusively on the conditional contract but instead use both incentives (the wage contract) and trust (the bonus), which more equally shares the risk from accidents across both the principal and the agent. Agents respond in kind. Our results seem to correspond with Fehr and Rockenbach (2003), who note that incentives perceived as fair do not seem to undermine reciprocity in the same way as ones considered unfair. Importantly, our study contributes to this literature by focusing on the design of appropriate incentive contracts in the presence of an inter-firm social dilemma.

Our new results on the wage-effort relationship add to the large gift exchange literature originating from Fehr et al. (1993), which demonstrates how reciprocity can elicit higher wages and effort than predicted by standard economic theory and therefore substantially improves welfare (see Charness and Kuhn, 2011, for a survey).[5] Fehr et al. (2007) show, for example, that voluntary, non-binding bonus contracts can substantially outperform explicit incentive contracts.

---

[2] Other experiments study whether gift exchange is robust to weakening the link between agent effort and output such as occurs with team production or separation of ownership and control in the firm. In particular, Cobo-Reyes et al. (2017) show that gift exchange still occurs but is weaker with teams of workers, especially when production technology makes coordination among workers more difficult (e.g. such as when output depends on the minimum or maximum effort exerted in the team). On the other hand, Maximiano et al. (2013) show that gift exchange is robust to separation of ownership and control, in which the manager decides the wage but the owner is a residual claimant.

[3] Andreoni (2018) reports a similar finding where in a modified trust game, trustees who voluntarily choose to offer a nonbinding guarantee are in fact more trustworthy. However, in his experiment, the trustors do not trust enough to increase efficiency.

[4] In a setting with uncertain output and unobservable effort, Hoeppner et al. (2017) find that exogenous penalties are more effective than contractual penalties imposed by the principal because the latter evokes negative reciprocity. This essentially confirms findings in the gift exchange literature on the potential negative effects of explicit penalties, although in those experiments output is deterministic.

[5] This gift exchange literature has its origins in Akerlof (1982) and Akerlof and Yellen (1990).

In addition to embedding the agency relationship within a social-dilemma, our experiment differs from the gift exchange literature in another important dimension: the agent's effort is unobservable and has a stochastic impact on performance. In almost all previous experiments, the relationship between effort and output is deterministic, and even though effort is not contractible, it is observable. Recent exceptions include Douthit et al. (2012), Rubin and Sheremeta (2016), and Davis et al. (2017), who all find that a noisy relationship between effort and performance significantly reduces both effort and wages. Random shocks thus greatly reduce the role of reciprocity in promoting efficiency. These existing papers consider only non-contractual bonuses.[6] In addition to considering inter-firm social dilemmas, our experiment builds on this existing work by examining whether contingent wage contracts, either alone or combined with ex-post voluntary bonuses, can help solve the agency problem when the outcome is stochastic.

2. Theoretical Model

Consider a principal-agent setting where the agent's effort decision influences the probability that an accident occurs. Greater effort lowers the probability of an accident but is costly. If an accident does occur, the principal is liable and suffers a decrease in earnings (i.e. the principal is penalized for the accident, such as when the firm is faced with civil liability). The principal cannot observe the agent's effort, but can observe the accident. An accident harms not just the principal but other principals in the group; i.e. the accident imposes a negative externality. We assume that both the principal (she) and the agent (he) are risk neutral.[7]

In particular, let e be the effort level, where $e \in [0, \overline{e}]$, and the cost of effort is $C(e)$ where $C'(e) > 0$ and $C''(e) \geq 0$. The probability an accident occurs is independent across firms and is denoted by $p(e)$, where $p'(e) < 0$ and $p''(e) \geq 0$.[8] If an accident occurs, then the principal is fined F, and all principals in the group are damaged by amount d. Each group consists of n principals,

[6] In related work, both Corgnet and Hernán-González (2019) and Chowdury and Karakostas (2020) investigate the trade-off between risk and incentives in the classic principal-agent problem. Both experiments consider only linear wage contracts, with the former using a real effort task, and the latter chosen effort.

[7] This turns out to be a relatively innocuous assumption in the discretized implementation of our experiment, since numerical calculations reveal that even substantial deviations from risk neutrality do not change the Nash equilibrium in the Baseline due to the discrete effort choices available and rounding.

[8] In an alternative set up the accident realization could be the same across all firms, with the likelihood depending on all firms' actions. This could represent certain types of environmental issues such as climate change or damage to other common pool resources. Such a common risk leads to an additional type of externality and changes the nature of the social dilemma, and is an interesting issue to investigate in future research.

indexed by *i.* Let R be the principal's revenue, which is independent of the effort choice.[9]

Since the inter-firm social dilemma is essential to our setting, we keep the features of the externality constant and instead vary the potential contractual forms to investigate how agency risk affects behavior in this novel social dilemma setting. The model considers three types of contracts where the principal offers the agent a wage, desired effort level, and sometimes a promised (discretionary) bonus. In one case, only a fixed (unconditional) wage contract can be offered. In the other cases, the wage offer can be conditional on whether or not an accident occurs. For comparison purposes, we begin by examining the case where there is no agent and principals make the effort choice themselves.

In this Baseline situation, the principal makes the effort decision directly and there is no agent. The expected profit of principal i is:

$$\mathrm{E}\Pi_i = R - C(e_i) - p(e_i)F - d\sum_{j=1}^{n} p(e_j)$$

Assuming (initially) the standard benchmark of purely self-interested preferences, maximizing this payoff function with respect to e_i yields the following first derivative

$$-C'(e_i) - p'(e_i)F - dp'(e_i).$$

If $(F + d) > -C'(0)/p'(0)$ then we have an interior solution, and the principal's individual optimal choice in this Baseline situation, $\boldsymbol{e_i^*}$**,** is defined by:

$$-\boldsymbol{p}'(\boldsymbol{e_i^*})(\boldsymbol{F} + \boldsymbol{d}) = \boldsymbol{C}'(\boldsymbol{e_i^*}) \qquad \textbf{(1)}$$

This is intuitive: the principal equates the marginal benefit from choosing a higher effort level with the marginal cost, where the marginal benefit includes a lower chance of incurring both the fine and suffering damage. As is standard in this type of externality problem, she ignores the harm inflicted on other group members.

[9] In the typical gift exchange experiment the agent's effort choice *does* affect the principal's revenue (or profit). In fact, it is by observing their revenue that the principal can infer the agent's effort choice even though they may not directly observe effort or contract over it. In our design, however, the principal can never observe or infer the agent's effort choice, akin to a standard principal agent framework. While alternative specifications are possible, we chose the simplest one. Of course, the principal's profit is affected by the agent's effort choice through the likelihood of a damaging accident occurring.

The second order condition for a maximum is

$$-C''(e_i) - p''(e_i)F - dp''(e_i) < 0$$

which is satisfied as long as either *C''(e) > 0* or *p''(e) > 0* or both.

To find the social optimum, maximize the sum of expected profits over the *n* principals in a group:

$$Total\ Expected\ Profit = nR - \sum_{j=1}^{n} C(e_j) - (F + nd) \sum_{j=1}^{n} p(e_j)$$

Differentiating with respect to e_i and solving yields the following first order condition, where $\boldsymbol{e_i^{**}}$ is the socially optimal effort choice:

$$-\boldsymbol{p'(e_i^{**})(F + nd) = C'(e_i^{**})} \qquad \textbf{(2)}$$

Comparing (1) and (2) reveals, unsurprisingly, that effort is higher in the social optimum compared with the Baseline individual choice; i.e. $\boldsymbol{e_i^{**} > e_i^{*}}$. This is summarized in the following prediction.

Prediction 1 (Baseline): *The principal chooses an effort level that is lower than the socially optimal level. (Equivalently, accidents will occur more frequently than is socially optimal.)*

Next, consider the principal-agent setting where only a fixed wage (unconditional) contract can be offered. Each principal has an agent also indexed by i. The principal can pay a discretionary bonus, however, after learning whether an accident occurred. In the first stage, the principal offers a contract $(w, \hat{e}, \hat{b})$ to the agent, comprising of a fixed wage w, the desired effort level $\hat{e}$ and the suggested bonus $\hat{b}$. In the second stage, the agent chooses his effort level e, which costs him $C(e)$, and he receives wage w. In the final stage, an accident occurs with probability $p(e)$. The principal observes whether an accident has occurred but not the effort level chosen by the agent. The principal then chooses the actual bonus, *b*.

In this case, the expected profit of principal *i* is:

$$\text{E}\Pi_i = R - w_i - b_i - p(e_i)F - d \sum_{j=1}^{n} p(e_j)$$

The agent's earnings (ignoring the bonus) are independent of the outcome:

$$w_i - C(e_i)$$

Denote the effort level, wage and bonus, chosen with this unconditional, Bonus Only contract as e_i^B, w_i^B and b_i^B, respectively. Since effort is costly, the standard economic prediction is that the agent chooses the lowest possible effort.[10] Anticipating this, the principal chooses the lowest possible wage and bonus. This is summarized as follows.

Prediction 2a (Bonus Only with Standard Selfish Preferences): *The principal offers the minimum wage, the agent chooses the lowest effort possible, and the principal pays no bonus.*

In contrast to this stark prediction arising from standard, selfish preferences, the large gift exchange literature demonstrates the importance of reciprocity in this context. However, as described in the previous section, a noisy relationship between effort and outcomes may undermine the benefits of reciprocity. This leads to the following alternative prediction:

Prediction 2b (Bonus Only with Gift Exchange): *The principal offers more than the minimum wage available and due to positive reciprocity the agent responds with more than the minimal effort. The principal reciprocates by paying a positive bonus. Due to the non-observability of the agent's effort, however, any potential gift exchange effect is weak and effort remains below the optimal level in the Baseline.*

Next, consider the case where the principal offers a Conditional (state-contingent) contract comprising two wages, but cannot pay a discretionary bonus. In the first stage, the principal offers a contract $(w_X, w_Y, \tilde{e})$ to the agent, comprising of a wage to be paid if an accident occurs w_X, a wage to be paid if no accident occurs w_Y, and the desired effort level $\tilde{e}$. In the second stage, the agent chooses his effort level e, which costs him $C(e)$. In the final stage, an accident occurs with probability $p(e)$. The principal observes the outcome, but not the chosen effort level, and pays the appropriate wage rate.[11]

[10] Note that since the agent has no choice but to accept the contract that is offered, no participation constraint is required.

[11] The incentive payment could also, in principle, be conditioned on the outcome of other firms. We do not consider that case, however, since such conditional wages are rarely observed in practice.

In this case, the expected profit of principal *i* is:

$$\mathrm{E}\Pi_i = R - p(e_i)w_{Xi} - (1 - p(e_i))w_{Yi} - p(e_i)F - d\sum_{j=1}^{n} p(e_j) \qquad (3)$$

The principal anticipates that the agent's effort choice is influenced by the wages offered. In particular, the agent's expected earnings are:

$$p(e_i)w_{Xi} + \big(1 - p(e_i)\big)w_{Yi} - C(e_i)$$

Differentiating with respect to e_i yields the agent's first derivative:

$$-p'(e_i)(w_{Yi} - w_{Xi}) - C'(e_i).$$

If $(w_{Yi} - w_{Xi}) > -C'(0)/p'(0)$ then we have an interior solution, and the agent's optimal effort choice with a conditional contract $e_i^C(w_{Yi} - w_{Xi})$, satisfies:

$$-p'\big(e_i^C\big)(w_{Yi} - w_{Xi}) = C'\big(e_i^C\big) \qquad (4)$$

Otherwise, the agent chooses the minimum effort possible. From (4), it can be shown that e_i^C is increasing in $(w_Y - w_X)$. Since wages are costly to the principal, and only the wage differential across states influences the agent's decision, the principal will set $w_X = 0$.

Comparing equations (4) and (1) reveals that in order to fully align the incentives of the agent with the principal so that $e_i^C = e_i^*$, the wage differential offered in the contract must equal the total harm suffered by the principal from an accident; i.e. $w_{Yi} = (F + d)$. However, it is not optimal for the principal to offer such a contract. To see this, substitute $e_i^C(w_{Yi} - w_{Xi})$ into (3) and differentiate to obtain the following derivative:

$$\frac{\mathrm{dE}\Pi_i}{dw_{Yi}} = -\Big(1 - p\big(e_i^C\big)\Big) - p'(e)\frac{\partial e_i^C}{\partial w_{Yi}}(F + d - w_{Yi}) \qquad (5)$$

The first term shows how the expected profit changes directly with the no-accident wage w_Y. The second term reflects the indirect effect that changes in wages have on the probability of an accident. In particular, a higher wage incentivizes a greater effort level, which reduces the probability of an accident. This in turn increases the likelihood of paying w_Y while decreasing the likelihood of harm from an accident caused by the agent. If $w_{Yi} \geq (F + d)$ then both the second term and the whole condition are negative, which cannot be optimal. Rather, the optimal wage contract, w_{Yi}^{C} comes from setting (5) equal to zero.[12] This is summarized in the following prediction.

Prediction 3a (Conditional Contract with Standard Selfish Preferences): *The optimal conditional wage contract involves paying a zero wage in the case of an accident, and a positive wage in the case of no accident. However, as it is not optimal to fully incentivize the agent, effort is lower than in the Baseline, but higher than in the Bonus Only treatment.*

As in the case of the *Bonus Only* contract, based on the gift exchange literature we conjecture that principals may offer higher wages to induce reciprocal agents to exert greater effort leading to the following alternative prediction.

Prediction 3b (Conditional Contract with Gift Exchange): *The principal offers higher wages than the selfish equilibrium wage offer, and due to positive reciprocity the agent responds with more effort than predicted by the (selfish and myopic) best response.*

If the principal can condition the wage payment on the observable accident outcome, she can provide ex ante incentives for the agent to exert effort but this exposes her to risk ex post from other firms' accidents. Bonuses may be important, therefore, to facilitate reciprocity and support a gift exchange outcome with greater effort and compensation, while also helping to mitigate some risk. Therefore, we also consider a setting with discretionary bonuses and outcome-contingent wages, labeled *Conditional+Bonus*. In the first stage, the principal offers a contract $(w_X, w_Y, \tilde{e}, \tilde{b})$ to the agent, comprising of a wage to be paid if an accident occurs w_X, a wage to be paid if no accident occurs w_Y, the desired effort level $\tilde{e}$ and the suggested bonus $\tilde{b}$. In the second stage, the agent chooses his effort level e, which costs him $C(e)$. In the final stage, an accident occurs with probability *p(e).* The principal observes the outcome, but not the chosen effort level, and pays the appropriate wage rate. The principal then chooses the actual bonus, *b*, so her expected payoff is

[12] This assumes an interior solution, which requires that (5) is positive at $e = 0$, a condition that holds in the experiment.

$$E\Pi_i = R - p(e_i)w_{Xi} - (1 - p(e_i))w_{Yi} - b_i - p(e_i)F - d\sum_{j=1}^{n} p(e_j)$$

Since the addition of the bonus does not change any of the derivation of equations (4) and (5), the equilibrium effort and wage contract in this conditional wage contract with bonuses is the same as in the conditional setting without a bonus: $e_i^{CB} = e_i^C$ and $w_{Yi}^{CB} = w_{Yi}^C$. The actual bonus paid is zero. Gift exchange and reciprocal preferences could lead to greater wages and effort, and positive bonuses. These predictions are summarized as follows:

Prediction 4a (Conditional + Bonus Contract with Standard Selfish Preferences): *The outcome is the same as in Prediction 3a, and the principal pays no bonus.*

Prediction 4b (Conditional + Bonus Contract with Gift Exchange): *The principal offers higher wages than the selfish equilibrium wage offer, and due to positive reciprocity the agent responds with more effort than predicted by the (selfish and myopic) best response. Consequently, the principal may pay a positive bonus.*

Comparing the results from the four cases described above with standard selfish preferences, we have the following hypotheses.

Hypothesis 1a: Efforts across treatments will be ordered $\boldsymbol{e_i^{**}} > \boldsymbol{e_i^{*}} > \boldsymbol{e_i^{CB}} = \boldsymbol{e_i^{C}} > \boldsymbol{e_i^{B}}$

Theoretically, the selfish equilibrium Conditional wage contract can incentivize the agent (e_i^C, e_i^{CB}) to take more than the minimum amount of effort, although this falls below the level of effort the principal would choose directly in the Baseline (e_i^*), and the socially optimal level (e_i^{**}). In the absence of reciprocity, the Bonus Only contract should induce the lowest effort level (e_i^B). While reciprocity (gift exchange) can lead to effort above the minimum, it may not fully solve the incentive problem.

These treatment comparisons can equivalently be expressed in terms of the associated accident probabilities:

***Hypothesis 1b*:** Accident probabilities across treatments will be ordered $\boldsymbol{p_i^{**}} < \boldsymbol{p_i^{*}} < \boldsymbol{p_i^{CB}} = \boldsymbol{p_i^{C}} < \boldsymbol{p_i^{B}}$

According to standard theory, wages in the Bonus Only case should be the minimum possible, but will be higher in the Conditional treatments. Reciprocity may lead wages in the Bonus Only treatment to be above zero, and also increase wages in the Conditional treatments.

Hypothesis 2: Expected wages across treatments will be ordered $pw_{Xi}^{CB} + (1-p)w_{Yi}^{CB} = pw_{Xi}^{C} + (1-p)w_{Yi}^{C} > w_i^B$

The scope for wages and bonuses to incentivize the agent and increase potential gains in the agency relationship is limited when individuals have standard selfish preferences, as summarized above in Hypotheses 1 and 2. Reciprocity and gift exchange, however, can encourage mutually beneficial increases in wages, effort and bonuses that make both the principal and agent better off. We therefore conjecture that our data will reject these hypotheses in the direction of greater wages, bonuses and efforts. The key question is which lever—conditional wages, ex post bonuses, or their combination—is more effective at raising agent effort and efficiency.

3. Experimental Design

Our experiment involves four treatments that correspond to the model described above: *Baseline*, *Bonus Only* contract, *Conditional* contract, and *Conditional+Bonus* contract. We chose a simple parametrization consistent with the model. In particular, we use a linear cost of effort function, $C(e) = e$, a non-linear accident probability function, $p(e) = \frac{0.9}{0.6e+1}$, and set $F = 0$.[13] We also set $d = 15, n = 4$ and $R = 35$. Earnings, wages, bonuses and *d* and *R* are denoted in experimental dollars (E$) and were exchanged for U.S. dollars at a pre-announced rate at the end of the experiment session.

[13] As shown in equation (1), the principal cares about the total earnings reduction inflicted on them by an accident which is the sum of the accident damage and the fine. Setting the fine equal to zero reduces the complexity for the experimental subjects. This also implies that the costs of an accident are shared equally by all principals in the group, maintaining the salience of the social dilemma aspect of the environment when the agency problem is introduced. Setting the fine to zero also spreads apart the noncooperative (*e**) and cooperative (*e***) effort choices in the Baseline condition. A comparison of equations (1) and (2) shows that this difference is determined by the relative size of *F*+*d* and *F*+*nd*, which is greatest when *F*=0.

In the experiment, effort choices are integer values from 0 to 10, with the corresponding probability of an accident ranging from 0.90 to 0.13. In all agency treatments, wages and bonuses are restricted to be integer numbers between 0 and 35. To ensure that subjects understood the accident probabilities, the experiment instructions (see Appendix C) illustrated the likelihood of an accident as a draw from a bingo cage containing 100 balls with two different types of balls. The accident probability determined how many "bad outcome" balls are in the cage. A table on subjects' computers displayed all the probability values for each of the 11 possible effort choices.

These parameters imply an individual optimal effort level $e_i^* = 3$ and socially optimal individual effort of $e_i^{**} = 8$ in the *Baseline* treatment.[14] The optimal wage contract in the *Conditional* treatments for selfish preferences is $w_X = 0$ *and* $w_Y = 3$, which results in $e_i^{CB} = e_i^C = 1$. Predicted effort in the *Bonus Only* treatment is the minimum effort level of 0; i.e. $e_i^B = 0$.

We chose parameters to ensure a reasonably large separation between the individual and social optimum effort levels in the *Baseline*, and to generate a strong incentive to solve the social dilemma. If all principals in the group choose the Nash equilibrium level of effort of 3 in the *Baseline*, then firm profit is E$12.71, while if all choose the social optimal of 8, firm profit is E$17.69, a nearly 40% increase. On the other hand, the variance of profits is large, as each accident in the group reduces earnings by E$15.[15] Choosing a range of effort levels up to 10 ensures both optimal levels are interior solutions and that neither is a focal point.

Another objective of our parameter choices was to permit wage or bonus adjustments to achieve similar expected equilibrium payoffs so that principals and agents in the agency treatments could avoid high levels of earnings inequality. Expected equilibrium payoffs (under selfish preferences) are -E$0.06 for principals and E$0.31 for agents in the Conditional treatments.[16] There is, however, significant potential for efficiency enhancing gift exchange interactions. In particular, if all principals could induce their agents to choose an effort of 3, such as through a wage or wage+bonus payment of 12 when an accident is avoided, then expected payoffs increase to E$7.57 for the principal and E$5.14 for the agent, a 5000% increase in the total expected payoff

---

[14] These are the optimal discrete choices. The continuous individual optimal level of $e_i^* = 3.08$ can be found from equation (1) and the social optimum of $e_i^{**} = 7.82$ from equation (2).

[15] Since each principal's revenue *R*=35, profits are certainly negative if three or more accidents occur contemporaneously in a group.

[16] While these low wages and efforts could lead to occasional negative payoffs, realized wages and efforts substantially exceed equilibrium levels; consequently, negative payments were not common enough for any subject to have negative earnings across the experiment.

of the firm. These same, higher firm earnings could be distributed even more equally between the principal and agent using a higher total payment to the agent.

3.1 Treatments

Subjects participate in only one of the four treatments. The *Baseline* treatment has no agents, and principals make their own effort choices and then learn whether an accident occurred.[17] The three agency treatments, as described in the previous section, have three stages. Table 1 presents a timeline of the experiment and a brief summary of all the treatments. We discuss each of these stages in more detail below.

Table 1: Timeline and Treatment Summary

	Treatments			
	Baseline	Bonus Only	Conditional	Conditional+Bonus
Stage 1		Principal offers binding wage w, paid in all states, nonbinding desired effort and cheap talk bonus	Principal offers binding wage w_X and w_Y, paid for state *X* or *Y*, nonbinding desired effort	Principal offers binding wage w_X and w_Y, paid for state *X* or *Y*, nonbinding desired effort and cheap talk bonus
Stage 2	Principal chooses costly effort level	Agent chooses costly effort level (unobserved)	Agent chooses costly effort level (unobserved)	Agent chooses costly effort level (unobserved)
Stage 3	Accident (*X*) may occur	Accident (*X*) may occur, Principal chooses bonus	Accident (*X*) may occur	Accident (*X*) may occur, Principal chooses bonus
Additional Tasks	Risk elicitation task (one gamble chosen for payment); Social value orientation task (one choice chosen for payment); demographic survey (flat payment)			

Note: Stages 1 through 3 are repeated for 15 periods with fixed matching (5 periods randomly selected for payment).

In the first stage, the principal offers a contract to the agent. In the second stage, the agent chooses their effort level e, which costs them $C(e)$. In the final stage, an accident occurs with

[17] An additional baseline could be one where there is a passive agent, who just receives a wage. However, such an agent would not have any choice to make, and therefore no reason to be paid a wage. While some principals might still wish to transfer some earnings to this passive agent, the motivation to do so would be analogous to a dictator game allocation and would be unrelated to the agency problem that is the most novel aspect of our study. For this reason we omit this treatment.

probability $p(e)$. The principal observes whether an accident has occurred but not the effort level chosen by the agent. Afterwards, the principal pays the contracted wage and in two of the three agency treatments chooses the actual bonus, b. As documented in Table 1, the difference between the agency treatments is the type of wage contract that can be offered and whether a discretionary bonus is allowed.

In all treatments, principals are randomly placed into groups of four that remain fixed throughout the experiment. Any accident among the group members reduces the profit of each principal in the group by 15 (d). It is important to note that accidents harm (only) principals, so even in the agency treatments where each group consists of four pairs, accidents are equally harmful to the group. Subject roles in the agency treatments are randomly determined at the beginning of the experiment and remain fixed throughout, as are the pairings between principals and agents.[18] No communication between subjects is possible, other than the wage and bonus offers and payments within the principal-agent relationship shown in Table 1.

After all the agents in a group make their effort choices, individual (own accident) and group outcomes (number of accidents in the group) are reported to all group members (the agents and the principals) but the effort choices of agents remain private information and are never revealed to other group members. Principals then choose the amount of the actual bonus in the *Bonus Only* and *Conditional+Bonus* treatments. Each period concludes with an earnings screen, displaying to subjects their own earnings for that period.

The experiment used neutral framing, with the outcomes described as X and Y (rather than accident or not), and the roles as A and B (rather than principal and agent). We also do not use the term "damages" but instead describe this as an "earnings reduction."

3.2 Procedures

Sessions were conducted at the Vernon Smith Experimental Economics Laboratory at Purdue University, using z-Tree (Fischbacher, 2007). All 184 participating subjects were undergraduate students, broadly recruited across different disciplines at the university by email using ORSEE (Greiner, 2015). We conducted 9 separate sessions. In all sessions, participants were

[18] Such fixed roles and pairings are common in the experimental gift exchange literature to explore the development of relational contracting. Since the stage game Nash equilibrium is unique and the finite number of rounds (15) is announced on the first page of the instructions, the repeated play does not change the equilibrium predictions of the benchmark model based on standard, selfish preferences. Repeated interactions are more relevant for managerial relationships in practice, however, and they also increase the potential relevance of reciprocity and gift exchange.

divided into groups of 8 or 4 depending on the treatment. We collected seven independent groups in each of the *Bonus Only* and *Conditional+Bonus* treatments, and six independent groups in the *Conditional* and the *Baseline* treatments.

At the beginning of each experimental session, an experimenter read the instructions aloud while the subjects followed along on their own copy. The main social dilemma task was repeated for 15 periods. Five of these periods were randomly selected for payment at the end of the session, with experimental dollars (E$) converted to U.S. dollars at a pre-announced 6-to-1 conversion rate in the *Baseline* and 3-to-1 conversion rate in the agency treatments.[19] Prior to beginning the task, all subjects participated in a computerized quiz to check their understanding of the experimental instructions, earning $0.50 for each correctly answered question.

Following completion of the main experimental task, subjects undertake two additional paid tasks to measure risk preferences and social value orientation (SVO), plus a demographic questionnaire. To elicit risk preferences, we use a scaled version of the Eckel and Grossman (2008) lottery choice, where subjects choose one of five gambles to be played out. Each lottery has two possible outcomes, both with an equal chance of occurring. The first is a safe option and the other lotteries increase in terms of their risk. To measure SVO, we use the six primary slider items from Murphy et al. (2011) with one chosen for payment at the end of the experiment. This enables us to classify subjects into four types: altruists, prosocial, individualists and competitive.

Subjects are paid their earnings in the instructions comprehension quiz, earnings from five randomly selected periods in the main task, earnings in the risk and SVO preference tasks, and an additional payment for completing the post-experiment survey. Subjects' total earnings averaged US$27.00, with an interquartile range of $20.75 to $33.25. Sessions usually lasted around 80-90 minutes on average, including the time taken for instructions and payments.

4. Results

We present results organized around the experimental hypotheses and use empirical approaches that provide a clean test of these hypotheses. Section 4.1 considers the effort decisions and resulting accident probability in each of the treatments, and Section 4.2 examines the principals' wage offers and bonuses paid. Section 4.3 summarizes the overall performance comparison across

---

[19] Principals make greater profits in the *Baseline* treatment because they do not need to pay wages to an agent, and the differential exchange rates roughly equalize average earnings across the treatments.

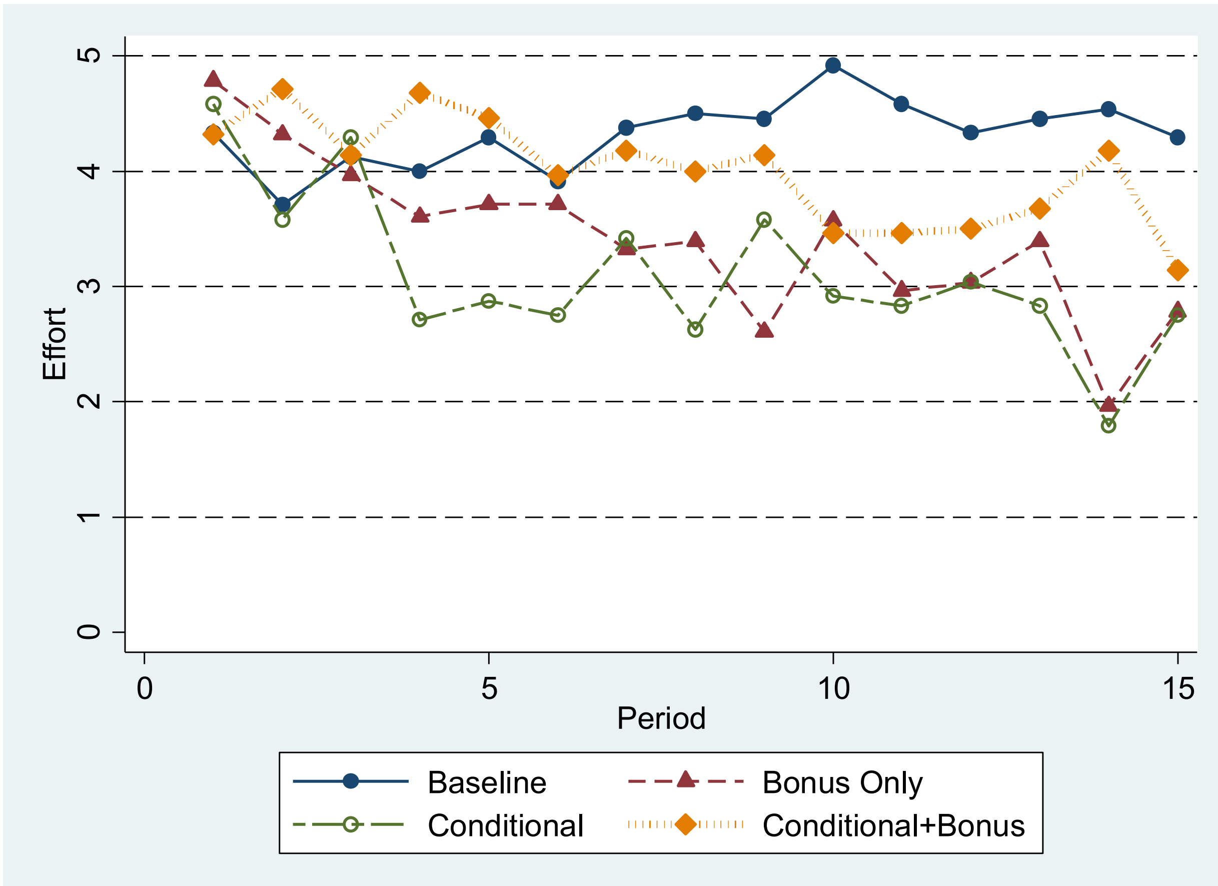
treatments in terms of earnings and damages. In each case, we present an overview of the data and report statistical tests focusing on non-parametric test statistics that use independent groups as the unit of observation. We also estimate panel data models with individual subjects representing random effects. These models cluster standard errors at the group level and include additional demographic, risk and social preference controls. All reported p-values are for two-sided tests.

4.1 Effort Choice

Figure 1 charts the evolution of mean effort decisions over the 15 periods, by treatment. While no systematic difference across treatments is evident in the first few periods, mean choices become more distinct as subjects gain experience. End period effects also appear in the three agency treatments. Our main interest is on stabilized behavior following some stationary repetition, rather than on learning; hence, while all the time series figures report data from all periods to document the broad patterns in the data, the treatment comparisons summarized below are based on the later periods 5-14, i.e. after the initial learning and adjustment phase and omitting the last period. Appendix A presents the descriptive statistics (Table A1) and the regression results (Tables A3-A4) for the entire sample, showing that the conclusions hold when considering all 15 periods.

Recall that Hypothesis 1 states that effort choices in the *Baseline* (no agent) treatment should be higher than the agency treatments, and should be lowest in the *Bonus Only* treatment where effort cannot be financially incentivized. Overall, effort choices in all treatments are on average between the individually optimal Baseline prediction (3) and the socially optimal level (8). Effort levels are significantly different from the social optimum in all treatments (sign rank test p-value < 0.05 in all treatments). Since subjects do not choose effort levels near the joint payoff maximizing level of 8, even for the *Baseline* treatment, behavior clearly reflects the tension between group- and individually-optimal behavior inherent in social dilemmas. The *Baseline* treatment average effort choice is significantly greater than the Nash prediction of 3 (sign rank test p-value=0.028). Average efforts in the *Conditional+Bonus* and *Conditional* treatments are significantly different from the equilibrium prediction of 1 (sign rank test p-values=0.018 and 0.028 respectively), and average effort in the *Bonus Only* treatment clearly exceeds 0 (p-value=0.028). These substantial deviations from the selfish preference model's equilibrium predictions indicates the important influence of reciprocity and gift exchange, and in what follows we focus our analysis on treatment comparisons.

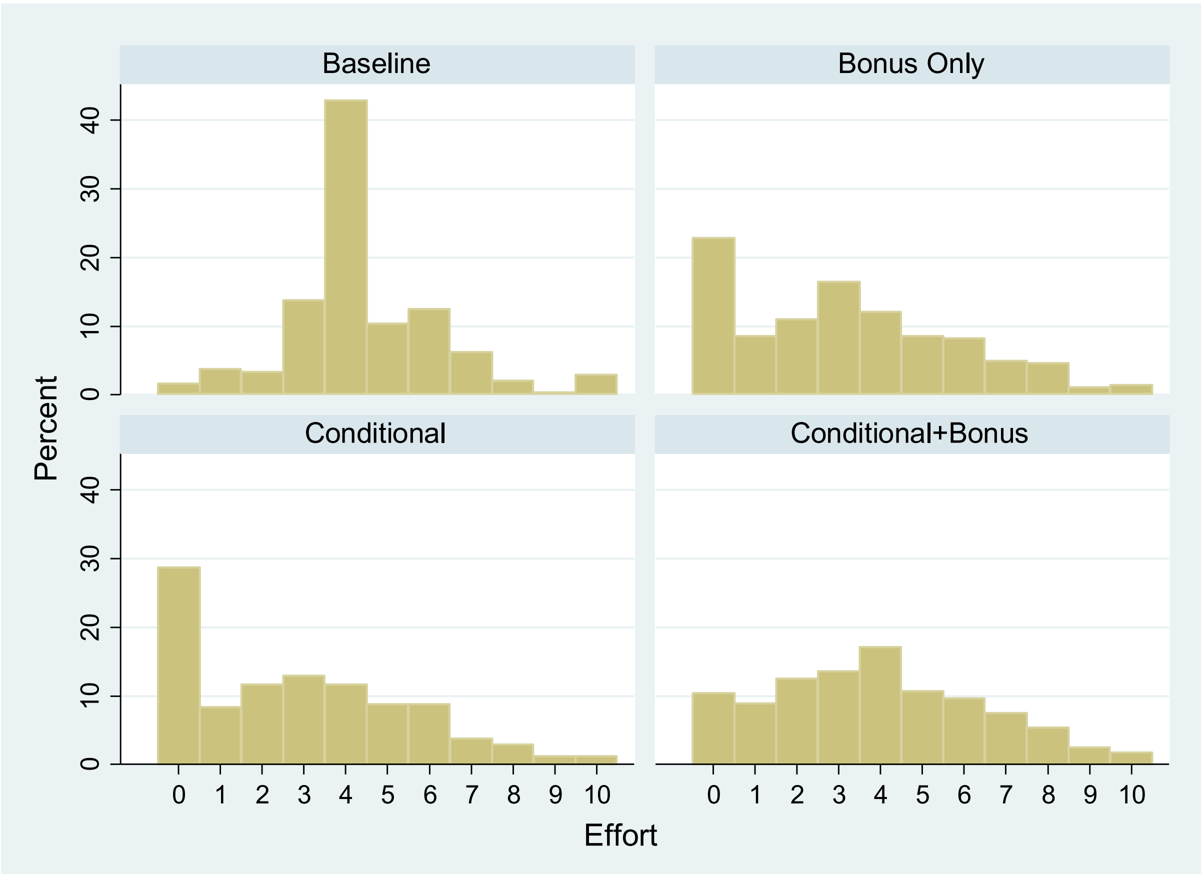
Figure 1: Average Effort across Treatments over Time



The overall mean effort obscures the fact that effort choices are widely dispersed, as illustrated in Figure 2. In all treatments, the choices vary across the entire feasible range [0, 10]. In the *Baseline* treatment, the strong mode on 4 is one unit above the Nash equilibrium, indicating a modest level of cooperation across the four unitary-actor "firms." In this treatment we observe complete free-riding, defined as (non-rationalizable) zero effort choices, only 1.67% of the time. By contrast, in the *Conditional+Bonus* treatment, the zero-effort frequency climbs to 10.36%, and in the *Bonus Only* (22.86%) and *Conditional* (28.75%) treatments agents make two to three times that number of zero effort choices.

Result 1: Effort choices are not statistically distinguishable in the *Baseline* and *Conditional+Bonus* wage treatments, and effort is significantly lower than the *Baseline* in the *Bonus Only* and *Conditional* wage treatments.

Figure 2: Distribution of Effort (Periods 5-14)



Support: A Kruskal-Wallis rank test rejects the hypothesis of equal efforts across all four treatments (p-value=0.021). Comparing effort levels in the *Baseline* treatment with the three agency treatments shows that the average effort exerted in the *Baseline* is significantly higher than the average effort in the *Bonus Only* treatment (4.44 versus 3.17; Wilcoxon ranksum test p-value=0.012) and the *Conditional* treatment (2.87; p-value=0.016), but not statistically different from the *Conditional+Bonus* treatment (4.44 versus 3.90; p-value=0.317). Effort in the *Conditional+Bonus* treatment is significantly greater than the *Bonus Only* treatment (3.90 versus 3.17, p-value=0.055) but is not significantly different from the *Conditional* treatment (p-value=0.133), due to higher across-session dispersion around the mean in the *Conditional* treatment. The *Bonus Only* and the *Conditional* treatments do not have significantly different effort levels (p-value=0.668). Overall, these findings are generally consistent with Hypothesis 1a, except for the insignificant difference between the *Baseline* and the *Conditional+Bonus* effort levels.

To investigate further how effort choices might depend on individual and group specific factors, Table 2 presents results from regression models of the effort exerted. Subjects in our experiment make multiple decisions in each session so these models include individual subject random effects. In addition, the errors are clustered at the group level to account for potential correlation of decisions within groups. We report five specifications. The first two use effort data from all four treatments. The coefficients reported in column 1 are from a parsimonious specification that only includes a time variable (period number) to capture the downward trend in effort, in addition to the treatment dummies. Column 2 includes demographic and other individual specific variables, such as risk preferences, gender, area of study, social value orientation, whether the subjects had an accident in the previous period interacted with their previous effort choice, and the total number of other accidents in the group.[20]

Both specifications provide similar results and are consistent with the non-parametric tests. Effort in the *Bonus Only* and *Conditional* treatments is significantly lower than in the *Baseline*, and no statistically significant difference exists between effort in the *Baseline* and the *Conditional+ Bonus* treatment. Wald tests of the equality of coefficients show that effort in the *Conditional+Bonus* treatment is significantly higher than in the *Bonus Only* and the *Conditional* treatment without the bonus (p-value<0.05 for all comparisons in both the first and second model specifications).[21]

Rather than simply including the accident outcome in the previous period in specification (2), unconditional on the amount of effort exerted, we constructed measures of "Good Luck" and "Bad Luck" to represent unlikely outcomes. The omitted case includes the most likely outcomes of having an accident following low effort or avoiding an accident following high effort, so these

[20] We create five dummies for the risk variable, with a choice of the safest lottery option as the reference dummy. The social value orientation scale shows that our subjects are mainly pro-social (105 subjects, 57.1%) or individualistic (78 subjects, 42.4%), with only one classified as competitive (0.5%). We create a dummy which is equal to one for subjects who are classified as pro-social.

[21] Subjects who choose a higher risk gamble as compared to the reference dummy, which is the safe choice, exert marginally significantly higher effort. Economics majors also exert marginally significantly greater effort, but other individual characteristics (e.g., gender, social value orientation) are not statistically significant.

Table 2: Random Effects Regression (Dependent Variable: Effort)

Dependent Variable: Effort	All treatments (1)	All treatments (2)	Bonus Only treatment (3)	Conditional treatment (4)	Conditional+ Bonus treatment (5)
Bonus Only	-1.270***	-1.305**			
	(0.416)	(0.508)			
Conditional	-1.571***	-1.676***			
	(0.497)	(0.460)			
Conditional+Bonus	-0.534	-0.652			
	(0.475)	(0.430)			
Period	-0.055*	-0.054*	-0.115	-0.031	0.003
	(0.029)	(0.030)	(0.078)	(0.025)	(0.065)
Good Luck (No previous		-0.760***	-0.695**	0.128	-1.577***
accident but Effort<3)		(0.275)	(0.271)	(0.229)	(0.458)
Bad Luck (Previous period		0.265	0.237**	0.387	0.489
accident but Effort>2)		(0.175)	(0.109)	(0.426)	(0.435)
Total other accidents in		0.060	0.036	0.259***	0.182
group (in previous period)		(0.065)	(0.095)	(0.074)	(0.156)
Desired effort			0.163***	0.180*	0.028
			(0.061)	(0.100)	(0.072)
Suggested bonus amount			0.071**		0.042*
			(0.028)		(0.023)
Wage offer			0.144***		
(unconditional)			(0.053)		
Wage offer for Y				0.272***	0.230***
(no accident)				(0.015)	(0.033)
Wage offer for X				-0.035	-0.089***
(accident)				(0.027)	(0.025)
Bonus amount received			0.049*		0.052**
in previous period			(0.027)		(0.025)
Constant	4.963***	3.972***	-0.055	-2.438***	1.540
	(0.386)	(0.712)	(0.831)	(0.784)	(1.174)
Demographic, Risk and Social Preference Controls	No	Yes	Yes	Yes	Yes
Observations	1040	1040	280	240	280
Number of agents	104	104	28	24	28

Notes: Data from Periods 5-14. Robust Standard Errors in Parentheses clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1

luck dummies are intended to capture *unexpected* and lucky outcomes deviating from expected performance in either direction. The "Good Luck" dummy variable captures the situation where the subject avoided an accident in the previous round even though he chose a low effort level (less than 3). The "Bad Luck" dummy variable is for cases where the subject experienced an accident in the previous round after selecting a high effort level (at least as high as 3). The significantly negative coefficient estimate shows that agents further reduce effort choice in the subsequent period following good luck, relative to the omitted case of the more likely accident outcomes.[22]The remaining three models in Table 2 consider the three agency treatments separately, in order to investigate the correlation of the two types of wage offers and previous and current suggested bonus amounts with effort. As frequently observed in the gift exchange literature, a higher wage offer by the principal is associated with a higher effort exerted by the agent, even in the *Bonus Only* treatment when the principal cannot condition the wage payment on the accident outcome. The principal's recommended effort and suggested bonus are also correlated with higher effort levels in this treatment.[23] For the *Conditional* treatments shown in columns (4) and (5), the wage offered for the good outcome (Y) has a strong positive impact on the effort choice. A lower wage offered for the accident outcome (X) is associated with a significant increase in agent effort only for the *Conditional+Bonus* treatment. In both of the treatments with bonuses, higher bonus payments in the previous period are associated with higher effort. This reflects reciprocity in the ongoing, repeated nature of the agency relationship.

Effort choices directly determine the likelihood of accidents, so treatment comparisons regarding accident probability generally parallel those for effort, although differences tend to be greater between the agency treatments. Appendix B provides detailed results, which show that consistent with Hypothesis 1b, the likelihood of accidents is significantly lower in the *Baseline* than in all three agency treatments. Accidents are also significantly lower in the *Conditional+Bonus* treatment than the other two agency treatments.

---

[22] Table A2 in Appendix A reports several alternative specifications, using different cutoffs (i.e., effort of 2 rather than 3 for good or bad luck), or simply a dummy variable for the previous period accident outcome independent of previous period effort, or a different omitted case (for an "expected" outcome of no accident when effort is at least 3). These alternative specifications do not change any of the treatment comparisons. They also show, consistent with the results in Table 2, that "good luck" in the previous period has a robust and negative influence on subsequent effort while previous "bad luck" has a weaker but consistently positive impact on later effort.

[23] This desired effort suggested by the principal is a nonbinding request and is largely ignored by the agent, and appears to devolve into mere cheap talk. The correlation between suggested and actual effort is weak, with correlation coefficients that range between 0.08 and 0.28 across the 3 agency treatments. Average desired effort ranges between 6 and 8. In contrast, as shown in the Baseline panel of Figure 2, when principals determine efforts directly, they choose efforts greater than 6 only 28 out of 240 times (12%). The modal desired effort in the agency treatments is the maximum (10), which principals essentially never choose in the Baseline. So the desired effort principals communicate to agents cannot be interpreted literally as what they wish agents to do.

4.2 Wage Offers and Bonuses

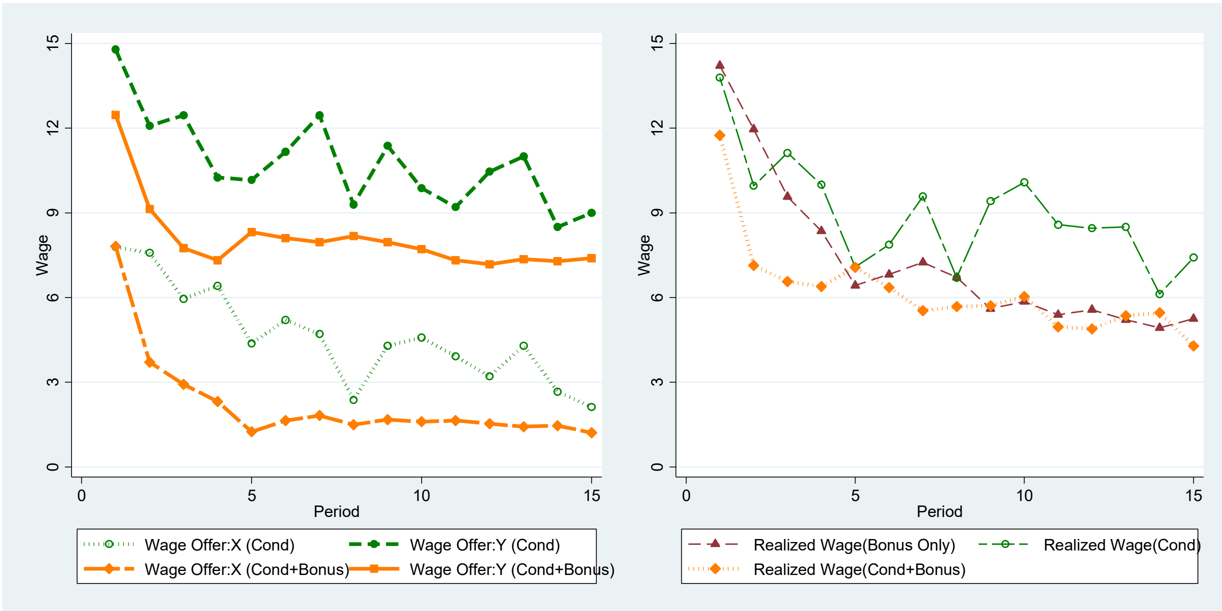
This section focuses on the payments made by principals to the agents, and therefore only considers the agency treatments. Principals in these treatments can suggest a certain amount of effort to the agent and except for the *Conditional* treatment they can also propose a suggested bonus. This is in addition to the binding wages they offer to the agent. We saw in the previous section that these decisions made by principals affect agents' chosen effort.

Result 2: Contrary to Hypothesis 2, average wages paid are not significantly different in the *Conditional+Bonus* treatment compared to the *Bonus Only* treatment. Realized wages are significantly higher in the *Conditional* treatment compared to the other two agency treatments. Principals do, however, pay much lower wages in the event of an accident when wages can be conditioned on the accident outcome.

Support: Figure 3 displays the time series of average wage offers. The left panel shows the average wage offers for each outcome in the two conditional treatments, while the right panel displays the average realized wages in the three agency treatments. The average realized wage in the *Conditional+Bonus* treatment is not significantly different from that in the *Bonus Only* treatment (5.71 versus 5.98; p-value=0.848). In the *Conditional* treatment the principal cannot pay a bonus, and realized wages are significantly greater than in the *Conditional+Bonus* treatment (8.24 versus 5.71; p-value=0.046) but are not significantly different from the *Bonus Only* (5.98) treatment (p-value=0.253).

Figure 4 provides more detail on the distribution of the different types of wage offers. About three-fourths of the wage offers for the accident outcome (X) are 0 in the *Conditional+Bonus* treatment, while less than half (44%) of the offers for the accident outcome are 0 in the *Conditional* treatment. Principals offer a wide range of wages for the good (no-accident) outcome (Y) in these conditional treatments, although about 70 to 85 percent are lower than the minimum offer of 12 needed to induce a self-regarding agent to choose an effort of 3. Nevertheless, as also clearly shown in the left panel of Figure 3, principals in the two conditional treatments offer significantly lower wages for the accident outcome (3.96 in *Conditional* and 1.56 in *Conditional+Bonus*) than for the no-accident outcome (10.35 in *Conditional* and 7.74 in *Conditional+Bonus*) (sign rank test p-value=0.028 and 0.018, respectively).

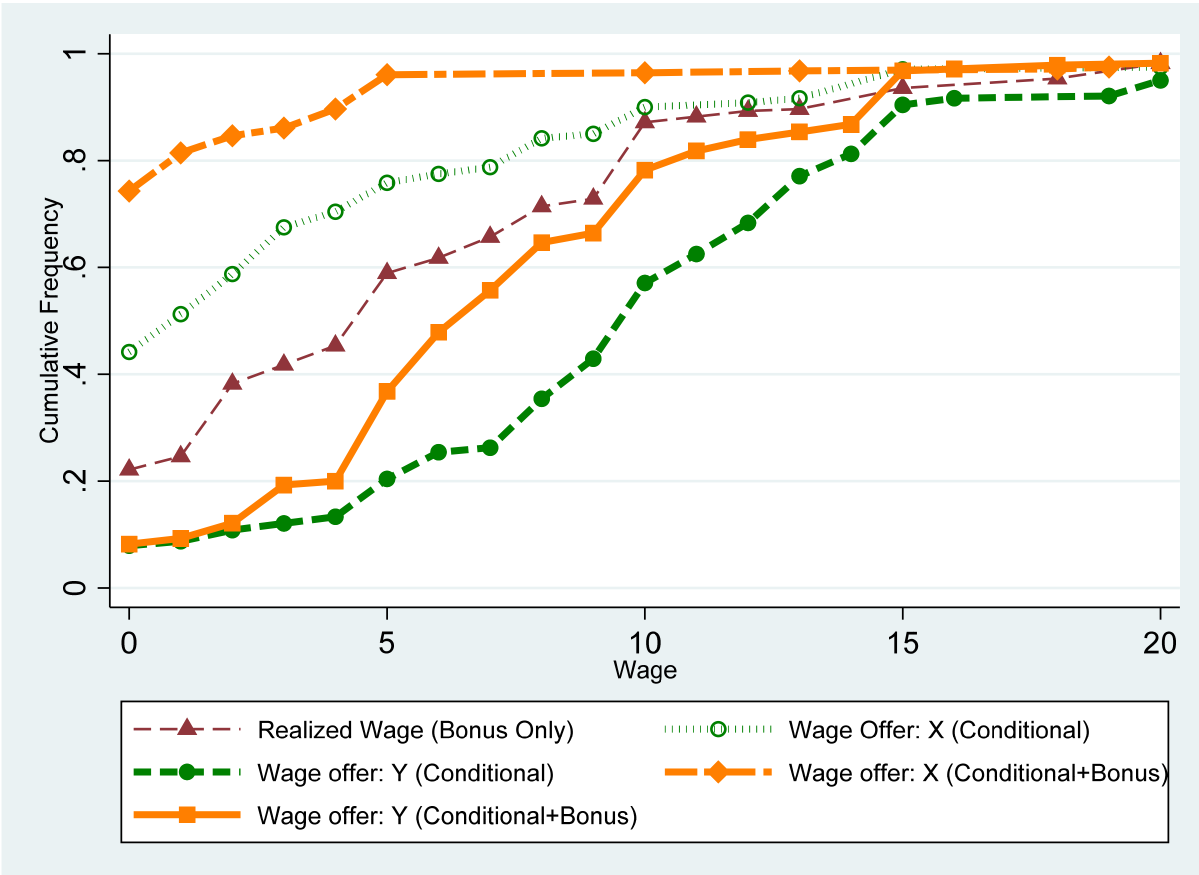
Figure 3: Average Realized Wages and Wage Offers over Time



The distribution of wages paid in the *Bonus-Only* treatment are also widely dispersed, like the wages offered for the no-accident outcome in the other treatments. As already shown in column (3) of Table 2, higher wage offers tend to elicit greater effort from the agent. The average wage in the *Bonus Only* treatment (5.98) is not significantly different from the no-accident (Y) wage offered in the *Conditional+Bonus* treatment (7.74; ranksum test p-value=0.338), but it is lower than the Y wage offered in the *Conditional* treatment (10.35; p-value=0.022). The average wage in the *Bonus Only* treatment is significantly greater than the wage offered for the accident outcome (X) in the *Conditional+Bonus* treatment (1.56, p-value=0.013) but not for the *Conditional* treatment (3.96, p-value=0.283).

Recall that under the standard model of selfish preferences, and considering the known endpoint of the repeated interaction, principals should not pay bonuses when they are feasible. Based on previous gift exchange experiments we hypothesized that bonuses might be paid due to reciprocity motivations. Especially since principals cannot condition wages on the accident outcome in the *Bonus Only* treatment, the motivation might be greater in this treatment.

Figure 4: Cumulative Distribution of Wage Offers (Periods 5-14)

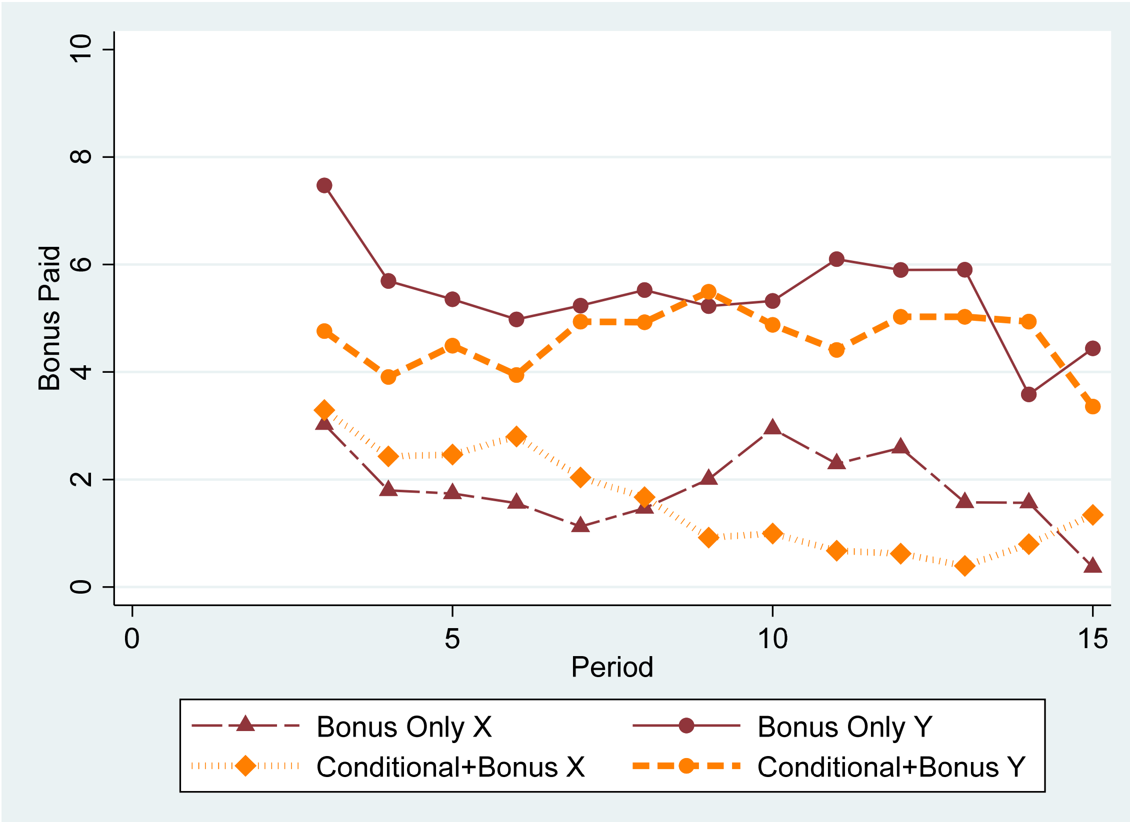


Result 3: Bonus payments are not significantly different in the two treatments where they are available.

Support: Figure 5 reports the average bonuses paid by the principal to the agent over the 15 periods. Because bonus choices are made subsequent to the accident realization, for both the *Bonus Only* and *Conditional+Bonus* treatments the figure reports separate time series for accident outcomes (labelled X) and the no-accident outcome (Y).[24] In both treatments, we observe positive bonuses, whereas under the standard model of selfish preferences the bonus is predicted to be zero. The bonus in the no-accident state is significantly higher than the bonus in the accident state in both *Conditional+Bonus* and *Bonus Only* treatments (sign rank test p-value=0.018 and 0.028 respectively).

[24] We report a moving average in the figure due to the high period-to-period variability in bonuses, arising in part from the small number of observations (separately for each outcome) underlying each average for individual periods.

Figure 5: Actual Average Bonus Paid over Time (3-Period Moving Average)



No significant differences exist across the agency treatments for the average bonuses paid (3.57 in *Conditional+Bonus* versus 3.51 in the *Bonus Only* treatment; ranksum p-value=0.565). The same conclusion holds when we separate the average bonuses for the good and bad accident outcome, shown in Table 3 (ranksum test p-value=0.701 for the no-accident outcome and 0.338 for the accident outcome).

Table 3: Average Bonus Paid for Agency Treatments, Conditional on Outcomes

Conditional+Bonus			**Bonus Only**		
Bonus Paid (Accident)	Bonus Paid (No Accident)	Suggested Bonus	Bonus Paid (Accident)	Bonus Paid (No Accident)	Suggested Bonus
1.09 (0.24)	4.92 (0.46)	8.05 (0.46)	1.59 (0.46)	5.00 (0.58)	12.65 (0.48)

Notes: Calculations based on periods 5-14. Standard errors shown in parentheses.

Table 4: Wages and Bonuses Paid by Principals

VARIABLES	All 3 Agency Treatments Wages Paid (1)	 Wages Paid (2)	Both Bonus Treatments Bonus Paid (3)	 Bonus Paid (4)
Bonus Only	0.271	0.750	-0.054	1.794*
	(1.317)	(1.136)	(1.189)	(1.058)
Conditional	2.535**	3.881***		
	(1.022)	(0.833)		
Period	-0.141**	-0.086	-0.173	-0.059
	(0.067)	(0.073)	(0.108)	(0.086)
Accident in this Period		-5.437***		-2.779***
		(0.913)		(0.554)
Accident in the Previous Period		-0.293		
		(0.448)		
Total other accidents (excluding own)				-1.675***
				(0.289)
Total other accidents previous Period (excl. own)		-0.183		
		(0.191)		
Suggested bonus amount				0.080**
				(0.039)
Constant	7.046***	6.422***	5.210***	0.765
	(0.937)	(2.098)	(1.478)	(2.985)
Demographic, Risk & Social Pref. Controls	No	Yes	No	Yes
Observations	800	800	560	560
Number of principals	80	80	56	56

Notes: Omitted case is the Conditional+Bonus treatment. Data from Periods 5-14. Robust standard errors in parentheses clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1

Although wages are a commitment made before the accident realization, it is possible that principals think of the wages plus the bonus paid as a combined compensation decision. This combination is greater than zero for the accident outcome (X) in both bonus treatments (2.02 for *Conditional+Bonus* and 6.13 for *Bonus Only*; sign rank test p-value=0.018 for both treatments). This combination is significantly greater in the *Bonus Only* than in the *Conditional+Bonus* treatment when the accident outcome occurs (ranksum test p-value=0.018). In contrast for the no accident outcome (Y), this combination is significantly greater in the *Conditional+Bonus* treatment (13.24) as compared to the *Bonus Only* treatment (12.09; ranksum test p-value=0.002).[25]

[25] Table 3 shows that the suggested bonuses are significantly above the actual bonuses given to the agents at the end of the period, in both treatments and irrespective of the accident outcomes (all p-values < 0.05). In the *Bonus Only*

Table 4 reports results from regressions of wages and bonuses paid, and confirms the patterns observed above. Wage payments are higher in the *Conditional* treatment, since no bonus compensation is possible in this treatment. Wages are significantly lower when an accident occurs, due to the principals' ability to pay different wages depending on outcomes in both conditional treatments. An accident by the agent and a higher number of accidents in the rest of the group both significantly reduce the bonus paid by the principal (column 4). Although not shown in the regression table, the individual-specific controls indicate that more pro-social principals offer a higher wage and also pay a higher bonus.

4.3 Earnings, Accidents and Overall Welfare

The wages, bonuses and effort choices translate into actual accidents and earnings, which importantly allow for overall welfare comparisons across treatments. Both the *Bonus Only* treatment, in which the principal must rely on nonbinding promises and ex-post bonus payments to incentivize effort, and the *Conditional* wage treatment, without any opportunities to pay bonuses, perform significantly worse than the *Baseline*. This is our final result:

Result 4: Actual accident frequency is greater and average earnings are lower in the *Bonus Only* and the *Conditional* treatment without bonuses than the *Baseline* treatment. There is no statistical difference between average earnings and accident frequency in the *Baseline* and the *Conditional+Bonus* treatments.

Support: Table 5 reports the average experimental dollar earnings and number of accidents per period for "firms" across the treatments. (For the agency treatments, each firm consists of one principal and one agent, while for the *Baseline* treatment each firm is a unitary actor.) If all four firms choose the socially optimal effort of 8, each would have an accident with probability 0.16 and average earnings would be 17.70 per firm. All treatments fall short of this performance benchmark. If instead all four firms choose the *Baseline* equilibrium effort of 3, each would have an accident with probability 0.32 and average earnings would be 12.71 per firm. While average earnings are not significantly different from this level for the *Baseline* treatment (sign rank test p-value=0.25) or the *Conditional+Bonus* treatment (p-value=0.13), they are significantly lower than

treatment, the difference from the suggested bonus is most pronounced. Nevertheless, column (3) of Table 2 indicates that the suggested bonus is associated with greater agent effort choices in the *Bonus Only* treatment.

12.71 for the *Bonus Only* treatment (p-value=0.028) and for the *Conditional* treatment (p-value=0.046).

Table 5: Average Earnings and Actual Accidents Per Firm, Per Period

	Baseline (No Agency)	Conditional	Bonus Only	Conditional+ Bonus
Earnings in E$	14.56 (1.69)	2.88 (2.21)	5.69 (1.78)	9.88 (2.41)
Realized Accident Freq.	0.267 (0.029)	0.488 (0.042)	0.436 (0.032)	0.354 (0.043)

Notes: Calculations based on periods 5-14. Standard errors shown in parentheses.

Comparisons between treatments reveal that realized earnings are significantly lower in the *Bonus Only* treatment (ranksum p-value=0.007) and the *Conditional* treatment (p-value=0.010) than the *Baseline*, but not significantly lower in the *Conditional+Bonus* treatment relative to the *Baseline* (p-value=0.133). The conditional wage treatments are marginally significantly different, with higher earnings when bonuses can be paid (p-value=0.063).[26] Parallel results obtain for the realized accident frequency, with the *Baseline* frequency significantly lower than the *Bonus Only* and *Conditional* treatments (both p-values<0.01) but not for the *Conditional+Bonus* treatment (p-value=0.114). A significantly greater accident frequency arises in the *Conditional* treatment when bonuses are not possible than when bonuses are possible in the *Conditional+Bonus* treatment (p-value=0.037). Accidents cause negative externalities and lead to a decrease in earnings for the whole group by an average of E$16 in the *Baseline*, E$21 in the *Conditional+Bonus*, E$26 in the *Bonus Only*, and E$29 in the *Conditional* treatment.[27]

4.4 Discussion of Findings

The fundamental innovation of our experiment is to embed the firm's agency problem in a new variant of a social dilemma, an inter-firm social dilemma. In this environment, effort cannot

[26] One reason that the *Conditional+Bonus* treatment has earnings that are not significantly different from any of the other treatments is the variance in earnings arising from the random accident outcomes. For expected earnings, constructed from the probability of suffering accidents, the *Conditional+Bonus* treatment is significantly different from all three other treatments at p-values no larger than 0.022.

[27] These total costs are simply calculated as the damage per firm (E$15) times the number of firms (4) times the average accidents realized.

be observed by the principal and effort does not deterministically map to outcomes. The findings suggest that conditional wage contracts outperform bonus-only, unconditional (gift exchange) contracts only when the conditional wages can be supplemented with bonus payments. Combining conditional wage contracts with bonuses incentivizes agents' effort choices (and the resulting probability of accidents) to approximate the performance of the *Baseline* condition, which does not have an agency feature.

Allowing principals to offer conditional wages as well as bonuses puts a heavy penalty on agents from having an accident, as principals offer significantly lower wages and pay very low bonuses in such cases. Agents respond to this wage differential, and they put in more effort to avoid an accident. Wage offers in the *Bonus Only* treatment do not appear to be high enough to encourage strong reciprocity. Wage realizations are however not different in the agency treatments when bonuses are feasible and no accident occurs, yet in the *Conditional+Bonus* treatment agent effort is higher and correspondingly accidents are less likely to occur than in the *Bonus Only* treatment. The wage-effort relationship therefore seems to be weaker when wages cannot be conditioned on effort or the stochastic (accident) outcome.

It is interesting to note principals use bonuses to incentivize agents in the *Conditional+Bonus* treatment, as documented in Table 3 and regression 5 of Table A4 in Appendix A. Our data help reconcile the use of bonuses with conditional wage contracts. The principals may use the bonus as a risk management strategy for both themselves and the agent, and also to help equalize their earnings. For example, the bonus can be a useful instrument to manage the group damage from accidents. Principals can mitigate against bad group outcomes (multiple accidents) by paying a lower conditional wage, but also limit the risk facing the agent by paying a bonus after outcomes are determined if they did not have an accident. By contrast, in the *Conditional* contract bonuses are not available, so the agent bears all the risk.

Another related explanation for the use of the bonus is principals' aversion to negative payoffs. For instance, if two outside firms have an accident she loses 30 (out of the starting revenue of 35), and so a no-accident wage greater than 5 would give her a negative payoff. To insure against this, the principal could offer a low no-accident wage (5 or less) in the conditional treatment but then pay a large bonus if less than two outside firms have an accident. Indeed, Figure 4 shows that nearly 40 percent of no-accident (outcome Y) wage offers in this treatment are 5 or less. In an additional regression of the bonus paid for only the *Conditional+Bonus* treatment, where we condition for the cases in which an accident did not occur, and control for the wages paid (since perhaps principals who already paid a generous no accident wage would not need to supplement this with a bonus), we find that the total other accidents in the group does have a strong negative impact on the bonus paid (p-value<0.001).[28]

Principals thus appear to use bonuses as a risk management strategy or a way to avoid negative payoffs. This can explain their use and effectiveness in the combined treatment, where accident-contingent wage contracts and ex-post bonuses are both possible.

5. Conclusion

This paper investigates principal-agent relationships and their impact when agents' unobservable actions affect others outside the relationship, leading to an externality and an inter-firm social dilemma. We design three agency treatments where principals can offer different kinds of incentives, some contractual and others that create motives for gift exchange. In the *Bonus* Only treatment, principals are limited to fixed (unconditional) wages and can only pay a bonus after the outcome of agent's effort is observed. In another treatment, *Conditional*, the principals can condition the agent's wage on observable outcomes but cannot pay bonuses. A third treatment, *Conditional+Bonus*, combines the conditional wage contract with a non-contractual bonus. We compare these three treatments to a *Baseline*, principals only, treatment.

Consistent with standard theoretical predictions, effort is significantly lower than the level required to achieve the social optimum even when there is no agency risk. Our findings from this Baseline treatment are also consistent with the large empirical literature on social dilemmas. Strikingly, effort choice is not significantly different across the *Baseline* and the *Conditional+Bonus* treatments, but is significantly lower in the *Bonus Only* and *Conditional* wage treatments. Correspondingly, the probability of having an accident is also higher in these two treatments.

Thus, principals use a combination of conditional wages and non-contractual bonuses to help solve the agency problem. Importantly, in our setting this implies that when it is possible to employ conditional contracts and bonuses, agency risk does not substantially worsen the inter-firm social dilemma. By contrast, the *Bonus Only* and *Conditional* treatments clearly exacerbate the

[28] This regression specification is similar to the one reported in Column 5 of Table A4, with the exception that this only includes data for when the firm does not have an accident. It also controls for the actual wage paid.

social dilemma, as they lack the combination of levers that seem necessary to incentivize the agent to exert greater effort. Since many social dilemmas have an agency feature, our findings highlight that designing appropriate agency contracts can be critical.

Our results contribute to a small literature establishing bounds on when the simple gift exchange relationship works well to incentivize agent's effort. Consistent with Rubin and Sheremeta (2016) and Davis et al. (2017), we find that in an environment with outcome uncertainty, relying solely on reciprocity is not enough to incentivize the agent. Similarly, explicit contracts that rely only on contractual wages contingent on the outcome also are not sufficient. For such challenging (and realistic) environments our experiment demonstrates that explicit (conditional wage) contracts along with bonuses perform better—and lead to the highest expected firm earnings of any of the agency treatments in this social dilemma setting. The combination of conditional wages and bonuses thus may be necessary to incentivize effort in the many problems, such as those relevant for financial regulation, that feature outcome uncertainty.

Our key objective of examining ways to reduce agency risk in the presence of a social dilemma required us to introduce both the inter-firm social dilemma and a probabilistic relationship between agent effort and (binary) performance, and the intent of our research is to identify effective contracts in this commonly observed setting. In future research it would be interesting to vary important features of the institutional environment to span the space between this new setting and the older experiments without a social dilemma to better understand the limitations and robustness of gift exchange to solve a wider range of agency problems. This could include, for example, treatments that held constant the type of feasible contract but varied the size of the externality and thus the importance of the social dilemma.

References

Abbink, Klaus, Jordi Brandts, Benedikt Herrmann and Henrik Orzen. 2010. "Intergroup Conflict and Intra-group Punishment in an Experimental Contest Game*." American Economic Review* 100, 420-447.

Akerlof, George. 1982. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics*, 97, 543-569.

Akerlof, George and Janet Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *Quarterly Journal of Economics*, 105, 255-283.

Andreoni, James. 2018. "Satisfaction Guaranteed: When Moral Hazard Meets Moral Preferences." *American Economic Journal: Microeconomics*, 10(4), 159-189.

Charness, Gary, Luca Rigotti and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review* 97, 1340-1352.

Charness, Gary and Peter Kuhn. 2011. "Lab Labor: What can Labor Economists Learn from the Lab?" in: *Handbook of Labor Economics* (vol. 4A), O. Ashenfelter and D. Card (eds.). Elsevier (Amsterdam). 229-330.

Chowdhury, Subhasish M. and Alexandros Karakostas. 2020. "An Experimental Investigation of the 'Tenuous Trade-off' Between Risk and Incentives in Organizations." *Theory and Decision*, 88, 153-190.

Claessens Stijn and Laura Kodres. 2014. "The Regulatory Responses to the Global Financial Crisis: Some Uncomfortable Questions" IMF Working paper Series, WP/14/46, Research Department and Institute for Capacity Development.

Cobo-Reyes, Ramón, Juan A. Lacomba, Francisco Lagos, and Dan Levin. 2017. "The Effect of Production Technology on Trust and Reciprocity in Principal-Agent Relationships with Team Production." *Journal of Economic Behavior and Organization*, 137, 324-338.

Corgnet, Brice and Roberto Hernán-González. 2019. "Revisiting the Trade-off Between Risk and Incentives: The Shocking Effect of Random Shocks?" *Management Science*, 65(3), 1096-1114.

d'Adda, Giovanna, Donja Darai, Nicola Pavanini, and Roberto A. Weber. 2017. "Do Leaders Affect Ethical Conduct?*" Journal of the European Economic Association,* 15(6), 1177–1213.

Davis, Brent J., Rudolf Kerschbamer, and Regine Oxel. 2017. "Is Reciprocity Really Outcome-Based? A Second Look at Gift-Exchange with Random Shocks." *Journal of the Economic Science Association*, 3, 149-160.

Dijk, Oege and Martin Holmén. 2017. "Charity, Incentives, and Performance." *Journal of Behavioral and Experimental Economic*s, 66, 119-128.

Douthit, Jeremy, Linwood Kearney and Douglas Stevens. 2012. "Can Agent Cheap Talk Mitigate Agency Problems in the Presence of a Noisy Performance Measure? An Experimental Test in a Single- and Multi-Period Setting." *Journal of Management Account Research*, 24, 135-158.

Eckel, Catherine and Phillip Grossman. 2008. "Forecasting Risk Attitudes: An Experimental Study using Actual and Forecast Gamble Choices." *Journal of Economic Behavior & Organization*, *68*(1), 1-17.

Erat, Sanjiv. 2013. "Avoided Lying: The Case of Delegated Deception." *Journal of Economic Behavior and Organization*, 93, 273-278.

Fehr, Ernst, Georg Kirchsteiger and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *The Quarterly Journal of Economics*, 108(2), 437-459.

Fehr, Ernst, Alexander Klein and Klaus Schmidt. 2007. "Fairness and Contract Design." *Econometrica*, 75(1), 121-154.

Fehr, Ernst and John List. 2004. "The Hidden Costs and Returns of Incentives-Trust and Trustworthiness among CEOs." *Journal of the European Economic Association*, 2(5), 743-771.

Fehr, Ernst and Bettina Rockenbach. 2003. "Detrimental Effects of Sanctions on Human Altruism." *Nature*, 422, 137-140.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economics Experiments." *Experimental Economics*, 10, 171-178.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association*, 1, 114-125.

Hoeppner, Sven, Lars Freund, and Ben Depoorter. 2017. "The Moral-Hazard Effect of Liquidated Damages: An Experiment on Contract Remedies." *Journal of Institutional and Theoretical Economics,* 173, 84–105.

Maximiano, Sandra, Randolph Sloof and Joep Sonnemans. 2013. "Gift Exchange and the Separation of Ownership and Control." *Games and Economic Behavior*, 77, 41-60.

Milkovich, George, Jerry Newman and Barry Gerhart. 2011. *Compensation* (vol. 10). McGraw-Hill/Irwin (New York).

Murphy, Ryan O., Kurt A. Ackermann and Michel J. J. Handgraaf. 2011. "Measuring Social Value Orientation." *Judgment and Decision Making*, 6(8), 771-781.

Rubin, Jared, and Roman Sheremeta. 2016. "Principal–Agent Settings with Random Shocks." *Management Science*, 62(4), 985-999.

Appendix A: Descriptive Statistics, Alternative Regression Specifications, and Results using the Full Data Set

The main text presents the entire time series for the variables of interest graphically, while the nonparametric statistics and regression results focus on treatment comparisons using data from periods 5-14. As explained in the main text, we do this because we are interested in examining behavior after the initial learning phase and omitting the last period. This Appendix first presents descriptive statistics for all decisions using the full data set and separately for the data from periods 5-14. It then presents estimates for the regression models using the entire data and shows that the primary conclusions are robust.

Table A1 presents the mean choices by treatment, separately for all periods and for periods 5-14. The across treatment patterns are similar for both panels.

Table A2 presents some alternative specifications for the pooled treatments regression of effort choices shown in Table 2. Columns (1) and (2) reproduce those same columns from Table 2. Column (3) replaces the good and bad luck dummy variables with simply a dummy variable to indicate whether the firm incurred an accident in the previous period, independent of the effort choice. Column (4) adds a dummy variable for the "expected" outcome of no accident when effort is at least 3, so the omitted case is now the expected outcome of having an accident with low effort. Columns (5) and (6) provide specifications analogous to columns (2) and (4), respectively, but using a different effort cutoff of 2 rather than 3 to define the good and bad luck outcomes. None of these alternative specifications meaningfully change the qualitative interpretation of the results.

Table A3 reports the effort decision regressions based on the entire dataset. The specifications in columns 1 to 5 follow those in Table 2 in the main body of the paper. As before we find support for Hypothesis 1a: effort choices are significantly greater in the *Baseline* treatment than in the *Bonus Only* and *Conditional* wage treatments, although they are not significantly different from the *Conditional+Bonus* treatment. The patterns observed relating to the relationship between wage offers and effort in the separate agency treatments retain their significance (columns 3 through 5). The variable capturing time, Period, is now statistically significant when pooling across treatments and for the *Conditional+Bonus* treatment separately, consistent with the initial downward trend in effort exhibited in Figure 1. The significance levels of some of the additional variables (Bad Luck, lagged bonuses received in the previous period) change somewhat relative to the main analysis using data from periods 5-14 reported in Table 2.

Table A4 replicates Results 2 and 3 in the paper. Wages paid are not statistically different in the *Conditional+Bonus* and *Bonus Only* treatments. Similarly, bonus payments are also not different across these two treatments, except marginally for specification (4) that includes additional control variables. Hence Hypotheses 2 is not supported by the data. Both wages paid and the bonuses exhibit a downward trend over time.

Table A1, panel a: Mean Choices by Treatment, All Periods

Treatment	Effort	Prob. of Accident	Realized Wage	Wage if Accident	Wage no Accident	Bonus (overall)	Bonus if Accident	Bonus no Accident
Baseline	4.322 (0.098)	0.283 (0.007)						
Bonus Only	3.410 (0.128)	0.415 (0.012)	7.277 (0.357)			3.962 (0.336)	1.610 (0.364)	5.675 (0.489)
Conditional	3. 111 (0.141)	0.456 (0.015)	8.981 (0.432)	4.633 (0.375)	10.806 (0.368)			
Conditional+Bonus	4.002 (0.126)	0.351 (0.011)	6.214 (0.295)	2.238 (0.257)	8.098 (0.257)	3.579 (0.255)	1.791 (0.328)	4.603 (0.340)

Note: Standard errors of the mean shown in parentheses.

Table A1, panel b: Mean Choices by Treatment, Periods 5-14 Only

Treatment	Effort	Prob. of Accident	Realized Wage	Wage if Accident	Wage no Accident	Bonus (overall)	Bonus if Accident	Bonus no Accident
Baseline	4.438 (0.118)	0.273 (0.007)						
Bonus Only	3.168 (0.155)	0.439 (0.016)	5.979 (0.359)			3.514 (0.396)	1.590 (0.463)	5.000 (0.579)
Conditional	2.867 (0.167)	0.477 (0.019)	8.242 (0.488)	3.963 (0.408)	10.350 (0.407)			
Conditional+Bonus	3.903 (0.151)	0.354 (0.013)	5.707 (0.357)	1.557 (0.289)	7.740 (0.314)	3.568 (0.329)	1.091 (0.244)	4.923 (0.462)

Note: Standard errors of the mean shown in parentheses.

Table A2: Random Effects Regression (Dependent Variable: Effort)

Dependent Variable: Effort	All treatments (1)	All treatments (2)	All treatments (3)	All treatments (4)	All treatments (5)	All treatments (6)
Bonus Only	-1.270***	-1.305**	-1.396***	-1.007**	-1.365***	-1.188**
	(0.416)	(0.508)	(0.521)	(0.496)	(0.510)	(0.502)
Conditional	-1.571***	-1.676***	-1.767***	-1.290***	-1.723***	-1.480***
	(0.497)	(0.460)	(0.469)	(0.430)	(0.466)	(0.449)
Conditional+Bonus	-0.534	-0.652	-0.737*	-0.413	-0.708	-0.590
	(0.475)	(0.430)	(0.442)	(0.395)	(0.448)	(0.438)
Period	-0.055*	-0.054*	-0.056*	-0.049*	-0.054*	-0.050*
	(0.029)	(0.030)	(0.030)	(0.027)	(0.030)	(0.027)
Good Luck (No previous accident but Effort<3)		-0.760***		-1.477***		
		(0.275)		(0.264)		
Bad Luck (Previous period accident but Effort>2)		0.265		0.080		
		(0.175)		(0.180)		
Accident previous period			-0.036			
			(0.150)			
Expected accident (Previous accident and Effort<3)				-1.119***		
				(0.225)		
Good Luck (No previous accident but Effort<2)					-0.805*	-1.292***
					(0.488)	(0.466)
Bad Luck (Previous period accident but Effort>1)					0.285*	0.158
					(0.163)	(0.159)
Expected accident (Previous accident and Effort<2)						-0.872***
						(0.220)
Total other accidents in group (in previous period)		0.060	0.045	0.087	0.046	0.057
		(0.065)	(0.066)	(0.070)	(0.063)	(0.063)
Constant	4.963***	3.972***	4.001***	4.271***	3.929***	4.097***
	(0.386)	(0.712)	(0.738)	(0.618)	(0.723)	(0.669)
Demographic, Risk and Social Preference Controls	No	Yes	Yes	Yes	Yes	Yes
Observations	1040	1040	1040	1040	1040	1040
Number of agents	104	104	104	104	104	104

Notes: Data from Periods 5-14. Robust Standard Errors in Parentheses clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Random Effects Regression (Dependent Variable: Effort)

Dependent Variable: Effort	All treatments (1)	All treatments (2)	Bonus Only treatment (3)	Conditional treatment (4)	Conditional+ Bonus treatment (5)
Bonus Only	-0.913***	-1.091***			
	(0.348)	(0.417)			
Conditional	-1.217***	-1.416***			
	(0.397)	(0.362)			
Conditional+Bonus	-0.320	-0.427			
	(0.402)	(0.354)			
Period	-0.072***	-0.062***	-0.101*	-0.037	-0.058**
	(0.022)	(0.023)	(0.052)	(0.024)	(0.027)
Good Luck (No previous accident but Effort<3)		-0.685***	-0.466**	0.069	-1.143**
		(0.241)	(0.228)	(0.287)	(0.479)
Bad Luck (Previous period accident but Effort>2)		0.301**	-0.040	0.130	0.520
		(0.152)	(0.269)	(0.292)	(0.442)
Total other accidents in group (in previous period)		0.063	0.133*	0.100	0.247
		(0.057)	(0.068)	(0.081)	(0.164)
Desired effort			0.152***	0.114	0.110
			(0.037)	(0.083)	(0.111)
Suggested bonus amount			0.067**		0.053**
			(0.034)		(0.024)
Wage offer (unconditional)			0.107**		
			(0.045)		
Wage offer for Y (no accident)				0.247***	0.178***
				(0.013)	(0.036)
Wage offer for X (accident)				0.007	-0.068**
				(0.022)	(0.027)
Bonus amount received in previous period			0.049***		0.058
			(0.018)		(0.037)
Constant	4.899***	4.155***	0.129	-2.173***	2.228**
	(0.276)	(0.599)	(0.525)	(0.410)	(0.913)
Demographic, Risk and Social Preference Controls	No	Yes	Yes	Yes	Yes
Observations	1,560	1,456	392	336	392
Number of agents	104	104	28	24	28

Notes: Data from all periods. Robust Standard Errors in Parentheses clustered at the group level.
*** p<0.01, ** p<0.05, * p<0.1

Table A4: Wages and Bonuses Paid by Principals

VARIABLES	Wages Paid (1)	Wages Paid (2)	Both Bonus Treatments Bonus Paid (3)	Bonus Paid (4)	Conditional +Bonus Paid (5)	Bonus Only Paid (6)
Bonus Only	1.062	1.280	0.383	1.526*		
	(1.211)	(1.161)	(0.990)	(0.921)		
Conditional	2.766***	3.989***				
	(1.011)	(0.785)				
Period	-0.366***	-0.183***	-0.206***	-0.114*	-0.088**	-0.160
	(0.055)	(0.062)	(0.064)	(0.061)	(0.040)	(0.126)
Accident in this Period		-5.383***		-2.779***	-2.117***	-3.457***
		(0.826)		(0.518)	(0.411)	(0.969)
Accident in the Previous Period		-0.499				
		(0.454)				
Total other accidents (excluding own)				-1.520***	-1.467***	-1.449***
				(0.175)	(0.211)	(0.259)
Total other accidents previous Period (excl. own)		-0.376*				
		(0.215)				
Suggested bonus amount				0.120***	0.049	0.156**
				(0.044)	(0.058)	(0.064)
Constant	9.142***	8.776***	5.230***	2.399	3.839***	2.321
	(0.772)	(2.328)	(0.896)	(2.167)	(0.668)	(2.482)
Demographic, Risk & Social Pref. Controls	No	Yes	No	Yes	Yes	Yes
Observations	1,200	1,120	840	840	420	420
Number of principals	80	80	56	56	28	28

Notes: Omitted case for columns (1) through (4) is the Conditional+Bonus treatment. Data from all periods. Robust standard errors in parentheses clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1

Appendix B: Accident Probability

This appendix presents results relating to Hypothesis 1b. Figure B1 plots the mean realized accident probabilities over time. In the *Baseline* treatment, the probability is relatively stable across periods, while the probabilities display an increasing time trend in the agency treatments particularly in the *Bonus Only* and *Conditional* treatments. The accident probability in the *Baseline* is below the theoretical prediction of 0.32 (sign rank p-value=0.046), consistent with the modal effort choice above the equilibrium noted in Section 4.1.

Result: Consistent with Hypothesis 1b, the likelihood of accidents is significantly lower in the *Baseline* than in all three agency treatments. Accidents are also significantly lower in the *Conditional+Bonus* treatment than the other two agency treatments.

Support: A Kruskal-Wallis test rejects that the data relating to probability of accidents are from the same population (p-value=0.001). Comparing between treatments using Wilcoxon tests confirm the differences evident from Figure B1. The probability of accidents is significantly lower in the *Baseline* than in the *Conditional+Bonus* (0.270 vs. 0.353; p-value=0.046), *Bonus Only* (0.270 vs. 0.438; p-value=0.003) and *Conditional* treatments (0.270 vs. 0.477; p-value=0.004). The probability of accidents is also lower in the *Conditional+Bonus* as compared to the *Bonus Only* (0.353 vs. 0.438; p-value=0.025) and *Conditional* treatments (0.353 vs. 0.477; p-value=0.022). Accident probabilities do not differ significantly between the *Conditional* and the *Bonus Only* treatments.

The result that the accident probability is significantly different between the *Baseline* and *Conditional+Bonus* treatments while the effort choices in these two treatments are not different is due to the significantly greater variance in effort choices for the *Conditional+Bonus* treatment.[29] While average effort is near 4 in both treatments, in the *Conditional+Bonus* treatment this mean arises through many offsetting very low and very high effort choices. By contrast, Figure 2 illustrates a strong mode of 4 for the *Baseline* treatment. Since the accident probability is a nonlinear function of the effort choice, the low effort choices frequently observed in the

[29] Although the central tendency of effort choices does not differ significantly between the *Baseline* and *Conditional+Bonus* treatments, Figure 2 suggests that effort variance is considerably greater in all of the agency treatments than in the *Baseline*. Differences in variance, first calculated within each independent session and then compared across treatments using Wilcoxon tests, are statistically significant between the *Baseline* and all agency treatments (p-value<0.05 for all cases). Similar tests show that effort variance is not significantly different across any of the three agency treatments.

Conditional+Bonus treatment translate to very high accident probabilities while the higher effort choices (above 4) only lead to relatively small reductions in accident probability. This leads to an overall larger increase in accident chances for the *Conditional+Bonus* treatment, relative to the modest reduction in mean effort.

Figure B1: Average Probability of Accidents across Treatments over Time

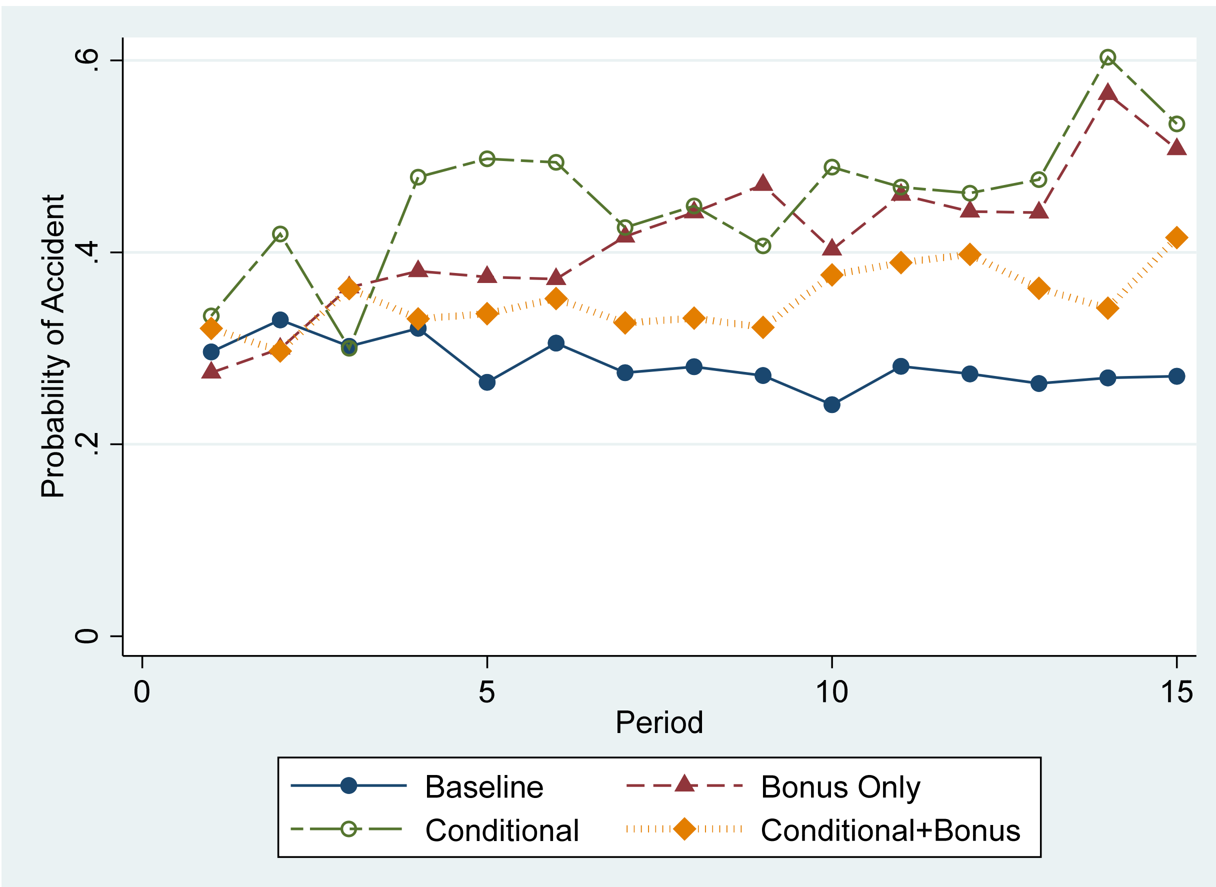


Table B1 reports random effects regressions for the probability of accidents and shows that the results described above supporting Hypothesis 1b are robust to the inclusion of additional control variables. The likelihood of accidents is significantly different across all three treatments, with the *Conditional* treatment having the highest accident likelihood and the *Baseline* the lowest. Table B2 indicates that these conclusions are robust to including all periods, and not only periods 5 through 14.

Table B1: Random Effects Regression (Dependent Variable: Probability of Accidents)

Dependent Variable: Probability of Accident		
	(1)	(2)
Bonus Only	0.166***	0.168***
	(0.026)	(0.035)
Conditional	0.204***	0.218***
	(0.041)	(0.042)
Conditional+Bonus	0.081***	0.097***
	(0.029)	(0.028)
Period	0.006**	0.006**
	(0.003)	(0.003)
Good Luck (No previous period accident but Effort<3)		0.068**
		(0.031)
Bad Luck (Previous period accident but Effort>2)		-0.013
		(0.013)
Total other accidents in group (in previous period)		-0.010
		(0.007)
Constant	0.214***	0.337***
	(0.030)	(0.073)
Demographic, Risk and Social Preference Controls	No	Yes
Observations	1040	1040
Number of agents	104	104

Notes: Data from Periods 5-14. Robust Standard Errors in Parentheses clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1 A test of the equality of treatment coefficients (not shown) indicates that accidents are significantly more likely in the Bonus Only and Conditional treatments compared to the Conditional+Bonus treatment (p-value<0.05 for all comparisons).

Table B2: Random Effects Regression (Dependent Variable: Probability of Accidents)

Dependent Variable: Probability of Accident		
	(1)	(2)
Bonus Only	0.131***	0.146***
	(0.022)	(0.031)
Conditional	0.173***	0.193***
	(0.032)	(0.033)
Conditional+Bonus	0.068***	0.084***
	(0.024)	(0.024)
Period	0.007***	0.007***
	(0.002)	(0.002)
Good Luck (No previous period accident		0.063**
but Effort<3)		(0.025)
Bad Luck (Previous period accident		-0.020
but Effort>2)		(0.012)
Total other accidents in group (in previous period)		-0.006
		(0.005)
Constant	0.225***	0.320***
	(0.021)	(0.057)
Demographic, Risk and Social Preference Controls	No	Yes
Observations	1,560	1,456
Number of agents	104	104

Notes: Data from all periods. Robust Standard Errors in Parentheses clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1 A test of the equality of treatment coefficients (not shown) indicates that accidents are significantly more likely in the Bonus Only and Conditional treatments compared to the Conditional+Bonus treatment (p-value<0.05 for all comparisons, except p-value=0.054 for Bonus Only compared to Conditional+Bonus in regression specification 2).

Appendix C: Experiment Instructions for *Conditional+Bonus* Treatment

Instructions **CB**

<u>Introduction</u>

This experiment is a study of group and individual decision making. The amount of money you earn depends partly on the decisions that you make and thus you should read the instructions carefully. The money you earn will be paid privately to you, in cash, at the end of the experiment. A research foundation has provided the funds for this study. Please put away your cell phones and other distracting devices for the duration of the experiment.

In this experiment, you will participate in three tasks. You will be given the instructions for the first task and after this task is completed, you will be given instructions for the next task. These tasks are independent so the decisions and payoffs from one do not affect the decisions and payoffs from the other.

Your earnings in this task are denominated in experimental dollars, which will be exchanged at a rate of 3 experimental dollars = 1 U.S. dollar at the end of the experiment.

<u>**Task 1**</u>

Overview

In Task 1 you will be in a group consisting of eight people, you and seven others. Your group members remain the same throughout task 1. The other people in your group are sitting in this room. You will make decisions privately, that is, without consulting other group members. Please do not attempt to communicate with other participants in the room during the experiment. If you have a question as we read through the instructions or any time during the experiment, please raise your hand and an experimenter will come by to answer it. At the end of these instructions you will take a computerized quiz and earn $0.50 for each correct answer you provide.

Task 1 is divided into 15 decision "periods." You will be paid based on your decision in five randomly chosen periods. Each decision you make is therefore important because it has a chance of determining the amount of money you earn.

At the beginning of this task each group of 8 will be divided into 4 pairs. Neither of you will ever know who the other person in your group is, but you will be paired with this same person for all

15 periods. The computer assigns each person a role, A or B. At the beginning of the task you know whether you are A or B. You will remain in the same role throughout the experiment.
Every period either outcome X or outcome Y will occur for each pair, and this is determined partly by chance and partly by decisions made by B. Regardless of whether the outcome is X or Y, A receives 35 experimental dollars each period. However if outcome X occurs for any of the pairs in your group, then every A in the group has their earnings for that period reduced by an amount -15.

The table below summarizes A's earnings:

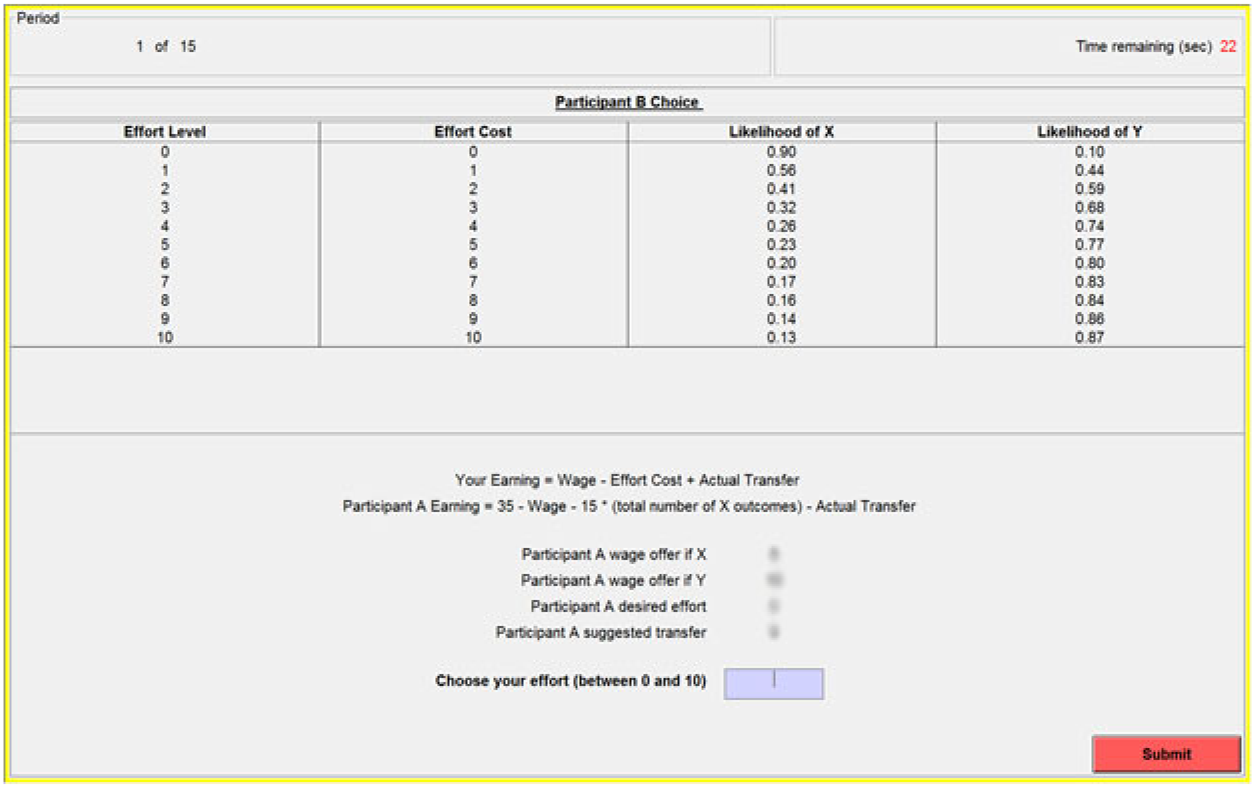
A's earnings	Number of group members with outcome X	Subtracted based on your group's outcome X
35	0	0
35	1	-15
35	2	-30
35	3	-45
35	4	-60

B's Decision

The decision made by B will affect the outcome in their pair. B's will need to choose an effort level (any integer number between 0 and 10) that will determine the likelihood of X or Y occurring in their two person pair. For each effort level chosen by B there is an associated cost of effort to increase the chances that outcome Y occurs. The cost of effort can be found in the table below. Your decision will be entered on a screen like the one shown below.

As you can see on the screen, choosing higher effort is more costly. The amounts shown in the second column indicate the amount that will be subtracted from your earnings based on your effort choice.

The right column on the screen shows how the chances of outcome Y increase for higher effort costs. You can think of the X and Y outcomes as occurring through random draws from bingo cages containing different numbers of X and Y balls. (The process will actually be conducted on the computer using a random number generator, but it is equivalent to this description based on bingo cages.)



Effort Level	Effort Cost	Likelihood of X	Likelihood of Y
0	0	0.90	0.10
1	1	0.56	0.44
2	2	0.41	0.59
3	3	0.32	0.68
4	4	0.26	0.74
5	5	0.23	0.77
6	6	0.20	0.80
7	7	0.17	0.83
8	8	0.16	0.84
9	9	0.14	0.86
10	10	0.13	0.87

For example, if you choose an effort level of 6, which has a cost of 6, the bingo cage would have 20 X balls and 80 Y balls. One ball will be drawn at random to determine the outcome, and so there is a 80 out of 100 chance (since there are 100 total balls) that the outcome is Y. If instead you choose an effort level of 1, which has a cost of 1, then the bingo cage would have 56 X balls and 44 Y balls. In this case, the chances of outcome Y would be 44 out of 100.

All random events are independent. This means that random outcomes occur as if new draws from new bingo cages occur every time a new random event is realized. A new cage and draw is used for every decision period and for every different individual.

A's decision

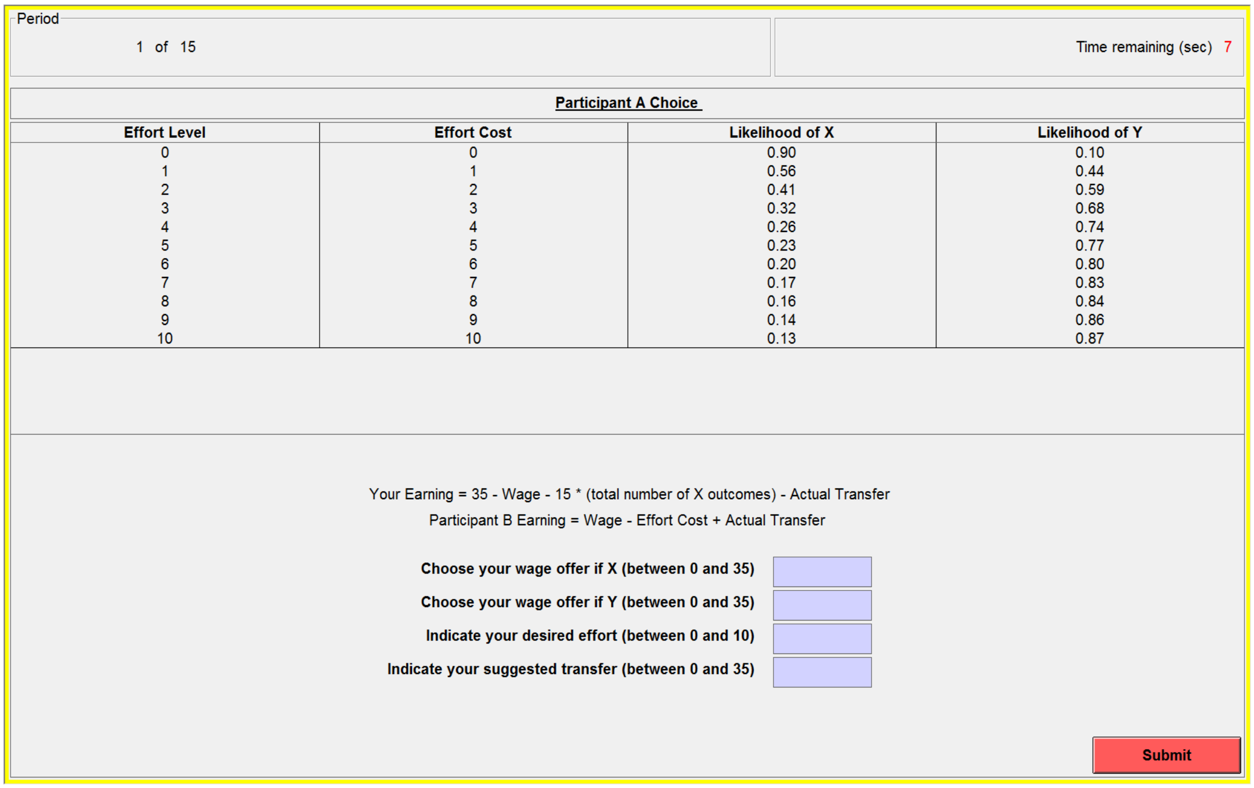
Each period, each A will need to make three decisions.

1. Choose two Wage levels to be paid to the B in their pair (any integer number between 0 and 35):
   - a wage level to be paid if the outcome X occurs, and
   - a wage level to be paid if the outcome Y occurs.

2. Recommend an effort level (any integer number between 0 and 10) for the B.

3. Suggest a Possible Transfer to the B in their pair (any integer number between 0 and 35). After the outcome is revealed in each period, A can decide how much to actually transfer to the B in their pair.

Before B’s choose their effort level, the computer will display the two wage levels, the recommended effort and the suggested transfer, to the B in your pair.


Period

1 of 15

Time remaining (sec) 7

Participant A Choice

Effort Level	Effort Cost	Likelihood of X	Likelihood of Y
0	0	0.90	0.10
1	1	0.56	0.44
2	2	0.41	0.59
3	3	0.32	0.68
4	4	0.26	0.74
5	5	0.23	0.77
6	6	0.20	0.80
7	7	0.17	0.83
8	8	0.16	0.84
9	9	0.14	0.86
10	10	0.13	0.87

Your Earning = 35 - Wage - 15 * (total number of X outcomes) - Actual Transfer

Participant B Earning = Wage - Effort Cost + Actual Transfer

Choose your wage offer if X (between 0 and 35)

Choose your wage offer if Y (between 0 and 35)

Indicate your desired effort (between 0 and 10)

Indicate your suggested transfer (between 0 and 35)

Submit

Earnings

After all the members of your group have made their effort decisions, the outcome for each pair (either X or Y) will be randomly determined by the computer as described above. Each pair sees their own outcome. A’s see the outcome but not the effort choice of B’s. After observing their outcome, including the total number of X outcomes in the group, A’s will decide the actual transfer to be given to the B in their pair.

Every period, A’s earnings depend on the choices of all the B’s in the group, as well as the outcome for their pair, as follows:

Earnings if the outcome is X = 35 - wage cost for outcome X – (15*number of X outcomes in my group) – actual transfer

Earnings if the outcome is Y = 35 - wage cost for outcome Y – (15*number of X outcomes in my group) – actual transfer

Every period, B's earnings will depend on the outcome for their pair as follows:

Earnings if the outcome is X = Wage for outcome X – cost of effort + actual transfer

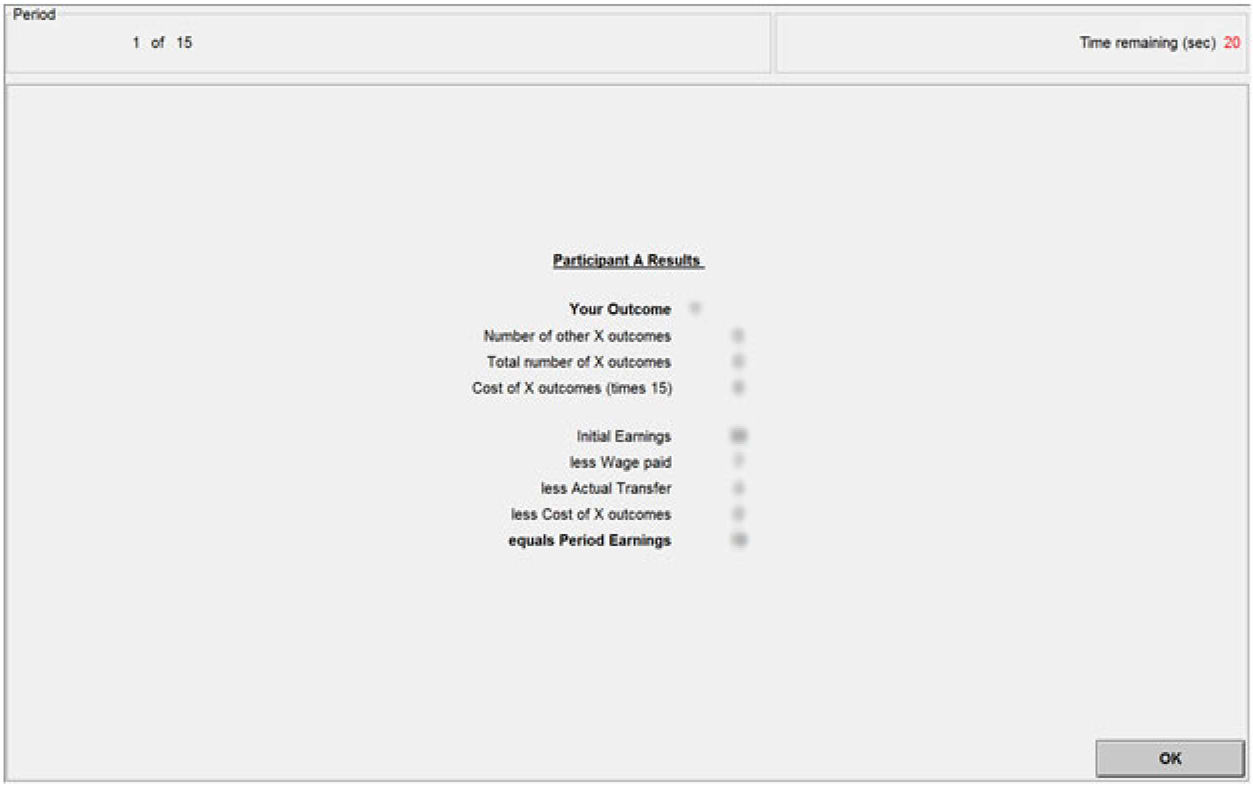
Earnings if the outcome is Y = Wage for outcome Y – cost of effort + actual transfer

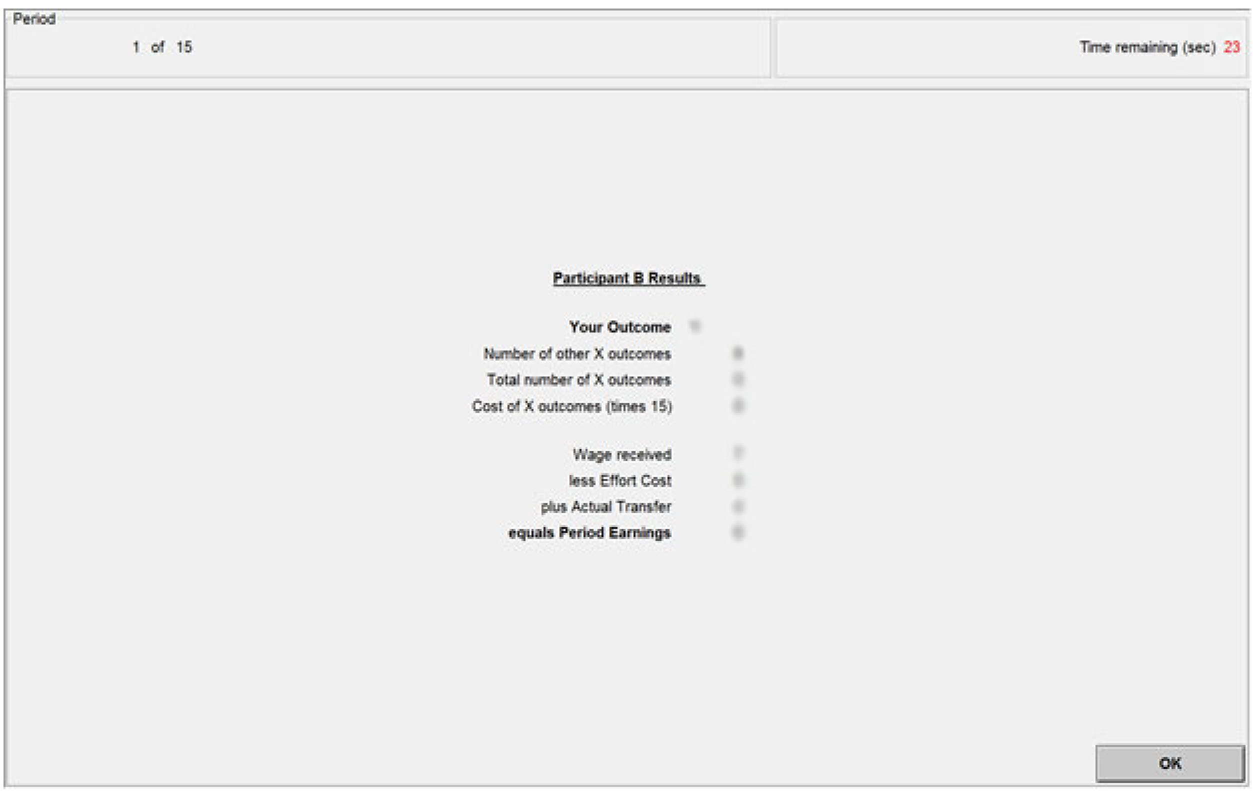
A results screen will display your earnings for the period, which will be paid in cash if this turns out to be one of the 5 periods drawn for payment. Which periods are drawn for payment is determined at the end of the experiment using a bingo cage at the front of the room. In addition to the wages and actual transfer paid, this results screen will display to both A and B the total number of X outcomes across all 4 pairs in the group.

Examples of the results screens for A and B are shown below.

Summary

- Throughout Task 1 you make decisions in a group of 8 that remains the same every period.
- Each group of 8 is divided into four pairs, with each pair consisting of an A and a B participant, whose roles remain fixed throughout the experiment.
- Individual pair members also remain fixed throughout the experiment.
- A needs to make three decisions: Wage (for each outcome); Recommended Effort; Suggested Transfer.
- After these are displayed to the B's in the respective pairs, B's choose their effort level keeping in mind the associated effort costs.
- The effort level by B determines the likelihood of outcomes X and Y for that pair.
- A does not observe the effort chosen by B, and only sees the outcome X or Y at the end of the period.
- If outcome X occurs, the earnings of all A's in the group are reduced by 15 for each X occurrence.
- After outcomes are revealed, A's can decide on the actual transfer level to B's.


Period
1 of 15
Time remaining (sec) 20
Participant A Results
Your Outcome
Number of other X outcomes
Total number of X outcomes
Cost of X outcomes (times 15)
Initial Earnings
less Wage paid
less Actual Transfer
less Cost of X outcomes
equals Period Earnings
OK


Period
1 of 15
Time remaining (sec) 23
Participant B Results
Your Outcome
Number of other X outcomes
Total number of X outcomes
Cost of X outcomes (times 15)
Wage received
less Effort Cost
plus Actual Transfer
equals Period Earnings
OK