The Effect of Contract Structure and Intentions on Agent Effort

Timothy Flannery and Stephen Roberts

April 16, 2019

Abstract

We design an experiment to test how contract structures and principal intentions affect agent decision-marking. Principals rank five contracts where a higher ranking increases the likelihood of a contract's implementation. In one treatment, those rankings are hidden from agents; in the other, agents observe rankings. Agent response to contract structure and monetary incentives largely conforms to our predictions: the vast majority respond favorably to contracts that theoretically encourage them to choose high effort, and agents also demonstrate a preference for monotonic contracts over non-monotonic contracts. Surprisingly, offering a flat contract that exactly compensates an agent for effort performs no better than offering the agent nothing. Principals do a poor job anticipating agent response to contracts, both when rankings are hidden and when they are observable. Although rankings only affect agent behavior in minor ways, the difference in rankings between treatments indicates principals choose rankings assuming they would significantly influence the decision of the agent.

Keywords: Principal-Agent Problems; Non-monotonicity; Experimental Economics; Contract Theory; Reciprocity; Intentions

1 Introduction

> The bottom line is that in many situations in which agents are making poor choices, the person who is misbehaving is often the principal, not the agent.
>
> Richard Thaler, *Misbehaving*, p. 190

The above quote come from a section of Thaler's book Misbehaving discussing the decision-making process of executives in the print media industry. It conveys the message that agents often make suboptimal decisions, not because they lack the skills to choose wisely, but because principals often fail to set up the right incentives for agents. This paper tests what elements of a contract agents value, and how well principals anticipate the responses of agents. We find that the majority of agents react to financial incentives, but some agents also demonstrate preferences for monotonic contract structures and reciprocity to generous contracts. Principal rankings substantially failed to adequately predict agent behavior. Principals appeared to significantly overweight the importance of reciprocity and intentions and underweight the importance of both financial incentives and contract monotonicity. As a result, the outcomes of the contracting situation in this paper mirrored the discussion from Thaler's discussion with executives: agents tended to respond both rationally and predictably, however principals failed to set proper incentives for those agents.

Our paper addresses several factors of contracting that either standard theory or behavioral theories have suggested are relevant to agent decision-making: financial compensation, contract shape (monotonic or non-monotonic), reciprocity (both to the intentions of the other party and to the realizations), and efficiency concerns. Our experiment tests agent responses and principal rankings for five different contracts, allowing us to compare agent preference over a variety of contract differences. Our experiment also includes two treatments, one where agents observe the rankings and another where they do not, to see the impact of principal intentions on agent behavior. Two contracts the principals rank have the same underlying monetary incentives for the agents and give principals the same expected payoff,[1] but one is monotonic in structure while the other is non-monotonic; therefore, this paper is the first to conclusively examine the effect of non-monotonicity on agent response.[2] These contracts could play a role in reducing risks in industries such as finance where more profit comes with increased risk, but such contracts only would work if agents still respond to them by selecting high effort. Such a contract may deter agent effort if agents perceive them as "unfair" since the principal earns more as profit increases but they earn less as profit

[1] Assuming agents choose the same option

[2] Flannery and Roberts 2018 also look at agent response to non-monotonic contracts; however, their study gave the principals a large amount of freedom in their choice of contract and agents rarely responded to a monotonic and non-monotonic contract with the same payoff. Brosig, Lukas, and Riechmann 2010 and Lukas 2007 also compare agent response to monotonic and non-monotonic contracts; however, they use contracts with different expected values to the agent and principal, which makes it unclear whether the different payoffs or the structure itself caused changes in agent behavior.

increases.

The contracts available for the principals to rank also allow us to explore how much non-monetary factors, such as efficiency or reciprocity, affect agent decision-making. Two contracts offer flat structures: the agent is paid the same regardless of outcome, but one of the contracts covers the agent's cost of effort and the other one does not. A difference between agent response to these contracts would indicate that agents are willing to promote efficiency, as long as it does not come at the cost of their own payoff. The fifth contract is a "generous" monotonic contract: agents would receive more selecting high effort than they would have received from the contract that paid them nothing, however the contract has higher expected monetary value to the agent for selecting low effort. This contract is of particular interest when compared across treatments: when agents can see the rankings of principals, ranking the "generous" contract more highly may lead to positive reciprocity on the part of agents. If principals believe agents will respond reciprocally, then there should be a noticeable difference in how the "generous" contract and the null contract are ranked between the two treatments.

Our results show evidence of impacts on agent response from both contract structure and reciprocity, however the most important factor in agent decision-making remains the monetary incentive. The two contracts that provided the strongest monetary incentive to choose high effort succeeded in eliciting the highest levels of high effort from agents. The different contract structures between these two contracts lead to a significant difference in agent response, with agents preferring monotonic over non-monotonic contract structures. The "generous" contract received a greater response from agents than other non-incentivizing contracts, providing support for the idea that agents wanted to act reciprocally. Agent response to the two contracts that paid the same regardless of outcome showed no evidence that agents were concerned with efficiency, even when they were compensated for cost of effort.

Our experiment also shows a gulf between principals and agents regarding intentions in contracting. In the observable treatment, principal rankings are displayed to the agents, allowing them to respond to each contract based on the principal's overall rankings. In the hidden treatment, agent response could not depend on principal rankings, because the rankings were not made visible to the agent. Agents selected nearly identical choices between the two treatments. Although agents responded more favorably to the monotonic contract the higher principals ranked it and less favorably to the non-monotonic the higher principals ranked it, changes in agent response were overall extremely minimal to contracts when principals ranked them differently. Principals, however, believed that agents would respond strongly to intentions, and substantially altered their rankings between the treatments. In the observable, this led to much worse outcomes for principals, as they ranked the "generous" contract highly despite only leading to high effort roughly one-third of the time, and the null contract earning the principal more than two other contracts. Principals who ignored the

observability of rankings and based their rankings off their own preferences when they tried the role of agents in practice rounds would have significantly outperformed those who based their rankings on their beliefs about agent concern for rankings.

2 Background

This experiment belongs to the class of papers testing principal-agent theory, as well as those examining the effect of intentions in labor markets. We designed the experiment to determine how agents respond to five different contracts that a principal can offer and to figure out which contracts principals want to offer. High effort is a best response for agents with two of these contracts: a monotonic contract and a non-monotonic contract. For the three remaining contracts, agents earn more by choosing low effort: a "generous" contract which more than compensates agents for the cost of high effort, a flat contract which exactly covers the agent's cost of effort, and the null contract which offers the agent nothing. Though no experimental papers have looked at how all of these different contract structures perform in a principal-agent environment, the vast literature on contracting and on other regarding preferences, such as a taste for fairness, provides insights on how subjects may respond to certain contracts, especially the generous non-incentivizing contract or the non-monotonic contract.

The experimental literature results on the principal-agent problem produced mixed results in how well they aligned with traditional theoretical predictions. Early experiments on the problem (Berg, Daley, Dickhaut, & O'Brien 1992; Epstein 1992) limited the contracting choices of the principal to a few options, and behavior by both the principal and agent roughly accorded with the subgame perfect Nash equilibrium. Later studies, such as Keser and Willinger (2000) and Flannery and Roberts (2018), find that subject behavior starkly differs from the subgame perfect equilibria when subjects have the opportunity to construct their own contracts instead of receiving just a few options. For example, Flannery and Roberts (2018) find that subjects often created contracts that may be perceived as "fair" but theoretically failed to incentivize effort. The current experiment limits the number of contracts to five, in order to see how agents respond to different contract structures and incentives. Restricting the contract options for principals may also lead to behavior more in line with the subgame perfect Nash equilibrium.

Several experiments, such as "gift-exchange" and "trust games" demonstrate that subjects often reciprocate generous behavior of others at the expense of earning more for themselves. In the gift-exchange laboratory experiments conducted by Fehr, Kirchsteiger, and Riedl (1993 and 1998), they find a positive correlation between effort level (or quality) and wages (or price) even though workers (or firms) have no theoretical incentive to choose higher effort levels (or quality levels). A field experiment (Gneezy and List 2006) shows that such payment schemes at least initially cause individuals to work harder but the effect disappears as time passes. In "trust games" (or "the investment game"), introduced by Berg, Dickhaut,

& McCabe (1995), the first mover can give money to a second mover that triples in amount, and the second mover has the opportunity to give the money back to the first mover. Even though the second mover has no incentive to give money back causing the first mover to have no incentive to give the second mover any money, first movers often give money and second movers often return it.[3] The choice for the principal and agent with the "generous" contract in this experiment is similar to the choices of the first in second movers in "trust games" and "gift-exchange games" because the principal is offering a contract expecting the agent to reciprocate kindly even though the agent has a theoretical incentive to choose low effort. However, the setup differs because in most of the "trust games" and "gift-exchange games" there is a continuum of choices while in this experiment the agent only has a binary choice, high or low effort.

Because supernormal profits can sometimes indicate an agent engaged in risky behavior (such as a financial advisor investing all of a client's money in one stock or even junk bonds), a few recent papers have explored the issue of non-monotonic contracts, those that actually may reward the agent less when the agent performs "too well". Recent research demonstrates that principals often fail to construct non-monotonic contracts when optimal, even against computer agents (Flannery and Roberts 2018). Since so few agents received non-monotonic contracts, the study provided no conclusive evidence on how agents respond to such contracts, particularly those that theoretically incentivize higher effort levels. This experiment provides the principal with both a non-monotonic contract and a monotonic contract with the exact same incentives and expected payoff. A difference in agent response to these two contracts indicates that agents have preference over contract structure.[4] Two other experiments also examine how agents respond to non-monotonic contracts compared to monotonic contracts, but the principal was only able to offer one of two contracts in these experiments (which may lead to experimenter demand effects), and they did not have the same expected payoffs causing concern that one may be favored over the other due to reciprocity instead of the structure of the contract[5] (Lukas 2007; Brosig, Lukas, and Riechmann 2010).

Economists sometimes discuss preferences for efficiency, and subjects often choose outcomes that are more efficient even at their own expense. Engelmann and Stroebel (2004) allow subjects to choose from a variety of payoff distributions and find that efficiency, selfishness, and maxmin preferences motivate their behavior more than fairness. Later research by Fehr, Naef, and Schmidt (2006) showed these results skew towards efficiency as the subject pool primary consists of economics and business majors who learn of the importance

[3]See Johnson and Mislin (2011) for a meta-analysis.

[4]Alternatively, if there is a difference in principal rankings between the contracts, principals believe that contract structure matters to agents.

[5]Additionally, their experiments use difficult probabilities for the subjects to comprehend such as 23% and 43% and have two periods in their design instead of just one as ours does. While the optimality of a non-monotonic contract in our experiment comes from a state with a payoff invariant to effort level, in both Lukas 2007 and Brosig et al 2010, it derives from correlation between the two periods.

of efficiency early in their education and may "self-select" into this major because of such preferences. Our experiment includes a flat contract which exactly compensates the agent for cost of effort but theoretically encourages the agent to choose low effort. If efficiency motivates agents, this type of contract is an inexpensive way for a firm to induce the agent to choose high effort.

Recent studies demonstrate that subjects care about intentions in addition to their actual compensation. Charness and Levine (2007) design an experiment where a medium wage can derive from a high wage offer with bad luck or a low wage offer with good luck. They find that subjects respond more favorably to a medium wage derived from high wage with bad luck than from an identical wage derived from a low wage with good luck, indicating that intentions matter and not just the underlying financial incentives. Falk, Fehr, and Schmidt (2000) also find that intentions matter in a moonlighting game where the first mover makes a choice in one treatment and the choice of the first mover is randomly determined in another. Our experiment also introduces a stochastic element to test whether intentions affect the decision of the agent. Unlike prior research, our experiment allows the first mover (the principal) to rank contracts where higher ranked contracts are given a higher probability of use in a stochastic distribution. In one treatment the agents observe the rankings while in the other they do not. Given the prior research, we expect the observed rankings to affect agent behavior. For example, if principals rated low paying contracts highly, it may increase the likelihood agents choose lower effort to all other contracts.

3 Experimental Design

The design of this experiment follows the framework of Flannery and Roberts (2018) with a few key differences. Like the Flannery and Roberts (2018) experiment, this one consists of a sequential game with a first and second mover (principal and agent respectively) with three possible outcomes: LOW, MEDIUM, or HIGH which pays the principal a revenue of 12, 36, and 48, respectively. The probability of the outcomes depends on the binary choice of the second mover (agent): high effort at a cost of 12 ECUs or low effort at a cost of zero ECUs. If an agent chooses low effort, the LOW outcome occurs three quarters of the time and the HIGH outcome occurs one quarter of the time. If the agent chooses high effort, the MEDIUM outcome occurs three quarters of the time while the HIGH outcome occurs one quarter of the time. The table below illustrates the effort, cost of effort, revenue from different outcomes, and the relationship between effort and outcomes.

Unlike the Flannery and Roberts (2018) paper, this paper restricts the principal (the first mover) who offers a contract over the state space (a payment for each possible outcome) to five possible contracts: (0,0,0); (0,16,20); (2,18,14); (6,14,22); and (12,12,12). The first number in each contract represents the payment from the principal (first mover) to the agent (second mover) for the LOW outcome, the second number for MEDIUM, and the third

Table 1: **Probabilities, Effort, and Revenue**

	% of *LOW* Outcome Rev. = 12	**% of *MED* Outcome Rev. = 36**	**% of *HIGH* Outcome Rev. = 48**
Effort = 0	3/4	0	1/4
Effort = 12	0	3/4	1/4

number for HIGH. Thus, if the contract (0,16,20) is used and the HIGH outcome occurs, then the principal earns the revenue from the HIGH outcome minus the payment to the agent, for a principal payoff of 28 in this case (48-20). Furthermore, instead of choosing one contract, the principal ranks the contracts 1,2,3,4 and 5 and the probability a given contract is implemented is 40% if ranked 1, 30% if ranked 2, 20% if ranked 3, 10% if ranked 4, and 0% if ranked 5. This design allows us to see how principals view each contract instead of just their top choice.

Once the principal ranks the contracts, the agent chooses a level of effort (high or low) for each possible contract before knowing which contract was implemented. Therefore, this experiment relies on the strategy method to learn how agents respond to each contract. Most of the literature confirms that this method typically yields similar results to the case where the agent knew exactly which contract was given.[6] Our experiment has two treatments: one where agents observe the rankings of the principal before choosing an effort level and one where the rankings of the principal remain hidden from the agent. In the treatment with hidden rankings, agents chose an effort level for all five possible contracts. In the treatment with observable rankings, we limited the agent to only choosing effort levels for outcomes that had a positive probability of occurring; thus, we omitted an effort level choice for the contract ranked fifth to avoid confusing subjects.

Agents receive an endowment of 20 ECUs in addition to the payment received from the principal (minus the cost of effort). Thus, if the agent chooses high effort, the contract (6,14,22) is selected, and the MEDIUM outcome occurs, the agent earns 22 (14[contract payment] + 20[initial endowment] - 12[effort cost]). The experimental design used this specific endowment for several reasons. Firstly, the endowment ensures that agents had positive earnings in the experiment. Secondly, the cost of high effort was a significant proportion of the endowment, which should cause the agents to think carefully about selecting high effort. Finally, the agent typically earns more than the principal with this particular endowment and set of contracts. Previous research (Keser and Willinger 2000; Flannery and Roberts 2018) shows that principals offer more to agents than expected, possibly due to a concern for fairness. When agents earn more than the principal, we expect the principal to behave more similar to how standard theory predicts, that is by maximizing one's own income.

[6]See Brandts and Charness (2011) for a survey of the literature on the strategy method.

3.1 Contract Structure and Agent Decision-making

Looking at the decision of the agent using backward induction, all contracts provide the agent more compensation for low effort except two contracts: (0,16,20) and (2,18,14). Since the HIGH outcome likelihood remains the same regardless of the agent's effort choice, an agent should select effort level based on the difference between the payment received for the MEDIUM outcome and the LOW outcome.[7] If the difference between them is sixteen or greater, the agent has an incentive to choose the high effort level. With a difference of exactly 16, the agent receives the same expected payment for each effort choice with (0,16,20) and (2,18,14); however, an agent with any of the following preferences would choose high effort: risk-aversion, inequity-aversion, or efficiency-maximizing. The design used the exact difference of 16 between the MEDIUM and LOW payments for two reasons. First, we did not want agents to receive too much more for high effort than low effort as we wanted to avoid making the decision trivial as that could elicit near one hundred percent high effort for both contracts, preventing us from observing how agents respond differ to different contract structures. Second, we wanted to ensure the principal would earn much more from offering these contracts if the agent chooses high effort compared to how much the principal earns for offering nothing with an agent response of low effort. The contracts (0,16,20) and (2,18,14) provide agents with the same expected payoff, and thus, if agents choose high effort more with one than the other, this can likely be attributed to the contract structure. The contract (2,18,14) has a non-monotonic structure since the agent earns more for the MEDIUM outcome than the HIGH outcome while the contract (0,16,20) has the more standard monotonic structure. Throughout the paper, we often refer to (2,18,14) as the non-monotonic contract and (0,16,20) as the theoretically incentivizing monotonic contract since agents have an incentive to choose high effort with this contract.

If agents behave as selfish profit maximizers, we expect low effort to be chosen with the contracts (0,0,0), (12,12,12), and (6,14,22). However, agents may be motivated by other factors beside monetary compensation. If agents primarily care about efficiency, then they will always choose high effort, regardless of the contract chosen as total surplus only depends on the action choice of the agent. If agents care about efficiency but want to be compensated for the cost of effort, they will choose high effort in response to all contracts but (0,0,0). The (12,12,12) contract exactly compensates agents for high effort but the monetary incentive to choose high effort exactly mirrors that of the (0,0,0) contract since the difference between the MEDIUM and LOW outcome payment is zero for both. Since companies often offer employees a flat salary, it is important to see the effect of higher payments for all outcomes (in this case by 12) on agent choice.

If agents are motivated by "generous" contracts, they may choose high effort with the contract (6,14,22) as it more than compensates an agent for effort. However, agents motivated solely by financial gain still have an incentive to choose low effort since the difference

[7] See the "Optimal Contracting" section Flannery and Roberts (2018) for more details.

between the MEDIUM and LOW payments is eight (less than the sixteen necessary to incentivize high effort).

If agents only care about the actual contracts and not intentions, then no difference should occur in the decision-making of the agents in the treatment where rankings are hidden and where rankings are observable. However, if agents care about the intentions of the principal, they may react with positive or negative reciprocity to different sets of rankings. For example, agents may choose low effort more frequently for all contract choices when the null contract is ranked one (highest) compared to when it was ranked five (lowest).

3.2 Contract Selection (Principal Decision-making)

The decision of the principal requires foresight as the principal must consider the actions of the agents before ranking contracts. If agents choose high effort for (0,16,20) and (2,18,14) but low for all others, the principal should rank the contracts as follows: 1 and 2. (0,16,20) and (2,18,14), 3. (0,0,0), 4. (6,14,22), and 5. (12,12,12). The principal earns an expected payment of 22 for (0,16,20) and (2,18,14), 21 for (0,0,0), 11 for (6,14,22), and 9 for (12,12,12) if agents act as if only motivated by financial gain. In each case, the principal makes less than the agent; therefore, principals with a concern for fairness should still choose this ranking (assuming they believe agents choose in a selfish manner).

Principals with the belief that agents are primarily motivated by efficiency concerns, should rank the contract (0,0,0) more highly. If principals believe agents are motivated by efficiency concerns but expect to be compensated for their cost of effort, then the contract (12,12,12) should be ranked highest. Finally, principals that think agents are motivated by "generous" contracts, would rank the contract (6,14,22) higher as it more than compensates the agent for cost of effort.

Principals may also believe that agents care about their intentions. When this is the case, rankings should differ between treatments. If principals rank the contracts differently between the treatments, then they believe their intentions play a role in the agent's selection of effort.

3.3 Details

The experiment was run at Missouri State University from February 26, 2018 to March 8, 2018 using SOPHIE software (Hendriks 2012). One hundred and thirty-eight students participated, 38 pairs in the observable ranking treatment and 31 pairs in the hidden ranking treatment. Each session lasted between 45 and 90 minutes with most sessions lasting roughly an hour. Funding was received through an internal research grant from Missouri State University. Each student received a $5 show-up fee[8] in addition to their earnings.

[8]Some students also received extra credit in a course for participating.

Agents earned on average $25.90 in the hidden ranking treatment and $27.08 in the observable ranking treatment, and principals earned on average $20.45 in the hidden ranking treatment and $19.34 in the observable ranking treatment.

Once students arrived, they received printed instructions on the experiment. A set of instructions for each treatment is available in Appendix B. After every student arrived, students logged into SOPHIE where they read through a more detailed set of instructions with an interposed quiz. Students then completed four practice rounds, two as each a principal and an agent. This allowed each student to understand the decision-making process of each role through experience. When all students finished the practice rounds, they started the eight paid rounds where they remained in the same role, either principal or agent, throughout the experiment.

4 Hypotheses

Our experimental design tests a number of hypotheses regarding the behavior of both principals and agents with different sets of contracts. Comparing between the two treatments allows us to determine how important knowing the principals' choices were to the agent, and also how important principals believed their rankings were. The first hypothesis deals how agents comparatively respond to each contract.

Hypothesis 1. *In the hidden treatment, the frequency of agent response of high effort will be (from highest to lowest): 1.(0,16,20) 2.(2,18,14) 3.(6,14,22) 4.(12,12,12) 5.(0,0,0).*

Hypothesis 1 is based off the expectation that monetary incentives will be the most powerful motivating factor in agent decisions. As a result, the contracts that incentivize high effort, (0,16,20) and (2,18,14) should cause the highest rate of high effort amount agents.

The difference between (0,16,20) and (2,18,14) is based off the prior research that has suggested agents prefer contract structures that are monotonic to structures that are non-monotonic. Contracts (0,16,20) and (2,18,14) both provide the same expected value to agents for selecting high effort (25 ECUs). Both contracts also provide 25 ECUs in expected value for choosing low effort. A difference in agent response between these two contracts then indicates a difference in preference regarding the structure of the contracts.

The differences between the three non-incentivizing contracts is based on belief that some agents will care about fairness or reciprocity. Agents concerned either about fairness or positive reciprocity may select high effort in response to contracts (6,14,22) and (12,12,12). The contract (6,14,22) in particular should incentivize high effort among agents who care about reciprocity, since the (6,14,22) contract gives the agent a guaranteed greater payoff when selecting high effort than selecting low effort with the null contract (0,0,0).

The hypothesis focuses on the hidden treatment because principals' intentions cannot affect the agents' choices in this treatment. Though we expect to see similar behavior from agents in the observable treatment, if agents do care about intentions, significant changes in agent response could result.

The second hypothesis also anticipates agent behavior. Agents are likely to respond more favorably to contracts that they perceive as being desirable if they are ranked highly, and respond more favorably to undesirable contracts if they are ranked lowly.

Hypothesis 2. *With observable rankings, agents will respond with high effort more often to (0,16,20) and (6,14,22) the higher they are ranked. Agents will respond with high effort more often to (2,18,14) and (0,0,0) the lower they are ranked.*

The contracts expected to be desirable to the agent are (0,16,20) and (6,14,22) because the contracts give a higher payoff when selecting high effort than the starting endowment, and they are monotonic in structure. The contract (2,18,14) may be considered undesirable because of its non-monotonic structure. If this contract is ranked lower, it might cause agents to reward the principal for this lower ranking. The contract (0,0,0) provides no incentive for the agent to select high effort, but one reason why the agent might take high effort is to reward a system of rankings from a principal. In this case, the agent may reward the principal for giving the contract (0,0,0) a low ranking.

In response to these expectations of agent behaviors, principals' rankings should be adjusted for the specific treatment. The treatment with observable rankings should cause principals to improve the rankings of favorable contracts such as (6,14,22) and decrease the rankings of unfavorable contracts such as (0,0,0).

Hypothesis 3. *When rankings are observable to the agent, principals will rank the contracts (6,14,22) and (12,12,12) higher, and the contract (0,0,0) lower, than when rankings are hidden from the agent.*

If principals believe that agents will respond according to how favorable they perceive the rankings to be, then contracts that pay the agent more such as (6,14,22) and (12,12,12) should be ranked higher when rankings are observable. On the other hand, the contract (0,0,0) is certainly viewed negatively by agents, so principals will likely feel the need to rank this much lower when rankings are observable.

The next hypothesis deals with the contracts that incentivize high effort, (0,16,20) and (2,18,14). Because these contracts incentivize high effort, they provide a higher expected value to the principal than any other contracts, if agents act as predicted.

Hypothesis 4. *Contracts (0,16,20) and (2,18,14) will be the two most highly rated contracts in both treatments.*

5 Results

In this section, we first consider general agent behavior and then test Hypotheses 1 and 2. Next, we turn to general principal behavior followed by tests of Hypotheses 3 and 4.

5.1 General Analysis of Agent Behavior

Table 2 below demonstrates the percentage of high effort responses to each contract in both the hidden and observable treatments. Therefore, in the first column with a contract, (0,16,20), agents responded to this contract with high effort 77% of the time in the hidden treatment and 76% of the time in the observable treatment.

Table 2: **Frequency of High Effort for Each Contract**

	(0,16,20)	**(2,18,14)**	**(6,14,22)**	**(12,12,12)**	**(0,0,0)**
Hidden	0.77	0.63	0.35	0.11	0.10
Observable	0.76	0.70	0.31	0.13	0.08

At a glance, the table displays no difference between the two treatments in terms of the aggregate amount of effort levels per contract. The largest difference occurs with the non-monotonic contract, with a 7% increase in high effort moving from the hidden to the observable treatment. Comparing the differences in response rates round by round, there is no significant difference in any of the eight rounds with this contract.[9]

5.2 Hypothesis 1

The top row in Table 2 illustrates that agents roughly behaved as predicted by Hypothesis 1. All of the statistics in this section focus on the difference between agents response in the treatment with the rankings hidden from the agents. In all eight rounds, a p-value of zero to six decimal places for the Cochran's Q test indicates that the response rate of agents was not simply random noise in any of the rounds. The (0,16,20) contract (the monotonic contract) elicited higher effort in every round than the other contracts with statistical significance in each round compared to the (6,14,22), (12,12,12), and (0,0,0) contracts.[10] As expected, agents responded more favorably to the monotonic contract (0,16,20) than the non-monotonic contract even though the monetary incentives for choosing high effort mirrored one another. Overall, agents responded with high effort 77% of the time with (0,16,20) compared to 63% of the time to (2,18,14); additionally, the contract (0,16,20) elicited a larger

[9]The largest difference occurs in the eighth round where twenty-six of thirty-five subjects chose high effort in response to the non-monotonic contract in the observable treatment while only seventeen of thirty-one responded with high effort in the hidden treatment; however, this differences is not even significant at the 10% level (2-sided Fisher's Exact Test, p=0.1241).

[10]With significance below the one percent level, using pairwise McNemar tests, with the exception of round 6 where the p-value comparing it to (6,14,22) is 0.0124, still significantly below 5%

proportion choosing high effort in all eight rounds.[11] Therefore, our results are the first to demonstrate agents prefer monotonic contracts to non-monotonic contracts when they provide the same monetary incentives. Though the non-monotonic contract performed worse than the monotonic incentivizing contract, it still elicited high effort 63% of the time, show that such contracts may have promise when they are the best theoretically available option.

The contract (2,18,14) exhibits a larger frequency of high effort significantly more than the (12,12,12), (0,0,0) and (6,14,22).[12] These results demonstrate the power of traditional theory in predicting agent behavior since (0,16,20) and (2,18,14) are the only contracts that fail to pay the agent more for low effort.

Unsurprisingly, agents choose high effort more often with the (6,14,22) contract than the (0,0,0) and (12,12,12) contracts.[13] Thus, the generous contract clearly elicits high effort more than the contracts (0,0,0) and (12,12,12), consistent with prior research that agents reward generosity; however, the contract also clearly underperformed the contracts (0,16,20) and (2,18,14) demonstrating that underlying monetary incentives also matter.

Though (0,16,20), (2,18,14) and (6,14,22) clearly caused agents to choose different effort levels among themselves and when compared to (0,0,0) and (12,12,12), the contracts (0,0,0) and (12,12,12) appear to elicit roughly the same level of effort with average percentage levels at 10% and 11% respectively. The contract (12,12,12) outperformed (0,0,0) (in terms of high effort proportion) in three rounds, the same in two rounds, and worse in three rounds. The difference in performance never differed significantly, even at the 10% level. Therefore, no evidence indicates that (12,12,12) performs better than (0,0,0) even though it offers the agent \$1.20 more and fully compensates the agents for their cost of effort! Offering agents more compensation without changing the underlying monetary incentives fails to encourage higher effort. Thus, employers who want to motivate agents to work harder should consider changing the contract incentives instead of simply paying more.

In the hidden treatment, nineteen agents chose high effort to at least one particular contract more than half the time. Twelve of them consistently selected high effort with the contracts that theoretically encourage high effort, (0,16,20) and (2,18,14). Five agents reciprocated by selecting high effort to generous contracts that more than compensated them for cost of effort: (6,18,22), (2,18,14), and (0,16,20). Finally, two agents only regularly chose

[11] However, the frequency of high effort only differed significantly between the contracts in five of eight rounds at the ten percent level, two at the five percent level, and once at the once percent level. They significantly differed in rounds 1,2,5,7, and 8 with p-values of 0.0578,0.0077,0.0334,0.0956, and 0.0578, respectively.

[12] With significance at the 1% level for (0,0,0) and (12,12,12) and at the 5% for (6,14,22) except once in round 2, where the difference between them only yields a p-value of 0.1336

[13] (6,14,22) significantly differs from the (0,0,0) and (12,12,12) contracts at the 5% level in all but one of eight rounds and two of eight rounds, respectively. In rounds 1 and 7, the difference between (6,14,22) and (12,12,12) gave a p-value of 0.1655 and 0.0956 (significant at the 10% level) while in round 8 the difference between (6,14,22) and (0,0,0) gave a p-value of 0.0956 (significant at the 10% level).

high effort to the contract (0,16,20), so these agents cared about both theoretical incentives and the structure of the contract itself. No agents frequently responded with high effort to any other sets of contracts. Hence, standard theory and, to a lesser degree, reciprocation describe agent motivations as opposed to agents motivated purely by efficiency or efficiency as long as it is a Pareto improvement over (0,0,0).

In the treatment with observable rankings, agents largely choose the same aggregate effort levels. Though not part of Hypothesis 1 due to the fact intention may have played a role in agent decision-making as well as sampling issues, the results are also consistent with prior research that the monotonic contract, (0,16,20), outperforms the non-monotonic contract, (2,18,14). Monetary incentives also matter since (0,16,20) and (2,18,14) encourage more effort than the others, and contract (6,14,22) yields a higher proportion of effort than (0,0,0) and (12,12,12). The difference between the contracts (12,12,12) and (0,0,0) is slightly greater with high effort percentages of 0.13 and 0.08; however, the difference is still small, and the amount of observations is smallest for (0,0,0) and (12,12,12) since so many principals ranked them worst.

Ultimately, the rankings of the contracts in the hidden treatment match exactly with Hypothesis 1. With the exception of the difference between the contracts (0,0,0) and (12,12,12), these ranking differences are statistically significant as well.

5.3 Hypothesis 2

Hypothesis 2 only concerns how agents responded to different rankings in the treatment with observable rankings. Table 3 below illustrates the proportion of high effort to each contract for a given rank in the observable treatment.

Table 3: **High Effort Proportion for Given Ranking**

	(0,16,20)	**(2,18,14)**	**(6,14,22)**	**(12,12,12)**	**(0,0,0)**
First	0.79	0.6	0.36	0.23	0.13
Second	0.79	0.66	0.28	0.08	0.11
Third	0.78	0.75	0.24	0.15	0.03
Fourth	0.69	0.79	0.31	0.11	0.05
Overall	**0.76**	**0.70**	**0.31**	**0.13**	**0.08**

According to Hypothesis 2, agents should respond more favorably to the contracts (0,16,20) and (6,14,22) when they are ranked higher since they have a structure agents are familiar with (monotonic) and they both more than compensate agents for cost of effort. The table shows no significant correlation using Kendall's Tau Rank Correlation Test between the ranking of (6,14,22) and (0,16,20) and an agent's response, but with only four rankings, the

test is quite weak.[14] Therefore, we compare how agents responded to each of these contracts when rated best and second best compared to when they were ranked third and fourth each round.[15] Agents responded to the (0,16,20) contract with more high effort in six of eight rounds in the observable treatment when ranked first (best) and second compared to when ranked third and fourth while they exerted an equal amount of high effort in round 2 and less high effort in round 7.[16] With the contract (6,14,22), in half the rounds agents responded with more effort when it was ranked first and second while in the other half with less high effort; thus, the response to the contract (6,14,22) appears to have no relationship with the rankings.

Only the non-monotonic contract appears to have a clear and significant (Two-sided Kendall's Tau Rank Correlation Test p-0.083) correlation with the ranking since agents respond more favorably to this contract when it is ranked lower. The proportion of high effort is larger in seven of eight rounds when the non-monotonic contract is ranked third and fourth compared to first and second.[17]

The results provide partial and weak evidence for Hypothesis 2. Lack of data caused us to draw no definitive conclusion on the relationship between the proportion of high effort and the ranking of the (0,0,0) contract. The ranking of the contract (6,14,22) appears to have no relationship with how agents respond to said contract. Nonetheless, there appears to be some evidence that a relationship exists between a contract's rank and the agent's response to that contract with the two contracts which incentivize effort: (2,18,14) and (0,16,20). Agents respond with more high effort to the monotonic contract the higher it is ranked while they respond less favorably to the contract the lower it is ranked while the opposite occurs with the non-monotonic contract. This is consistent with agents having a preference for monotonic contracts, even if a non-monotonic contract provides the same expected payoff and underlying incentives to choose high effort. Agents negatively reciprocate when the non-monotonic contract is ranked highly and positively reciprocate when a monotonic contract is ranked highly.

[14]With only four rankings, significance only happens when the relationship between the two variables is purely monotonic. The aggregate high effort proportion to the first (best) ranking was 0.787514 while to the second best ranking was 0.792683, so there is no significant relationship with ranking and high effort level with the (0,16,20) contract using Kendall's Tau Rank Correlation Test.

[15]We omit other comparisons such as first best ranking high effort proportions versus second, third, and fourth due a lack of observations.

[16]None of the differences were significant with the largest difference occurring in the third round only yielding a p-value of 0.1642 (Two-sided Fisher's Exact Test).

[17]However, as with the (0,16,20) contract, there is not a significant difference in any round with the largest difference occurring in the first round, giving a p-value of 0.1034 (Two-sided Fisher's Exact Test).

5.4 Analysis of Principal Behavior Overall

Table 4 displays the average ranking of each contract in both treatments of the principal. The Friedman Test demonstrates a significant difference between at least one contract and the rest in all rounds of the observable treatment (max p-value=0.0025).[18] The large disparity between the ranking of the (0,0,0) contract and the others is likely the primary cause of this. In the hidden treatment, the Friedman test only yields significance in the last three rounds at the 5% level. This significance likely occurred due to the decline in the ranking of the (12,12,12) contract over time.

Table 4: **Average Ranking for Each Contract**

	(0,16,20)	**(2,18,14)**	**(6,14,22)**	**(12,12,12)**	**(0,0,0)**
Hidden	2.90	2.58	2.84	3.46	3.23
Observable	2.73	2.70	2.45	3.23	3.9

The largest difference in averages between treatments occurs with the (0,0,0) contract, moving from 3.23 to 3.9. This change is consistent with Hypothesis 3 and unsurprising given that principals may not want agents to know they offered them (0,0,0) highly. The next largest difference in averages occurs with the (6,14,22) and (12,12,12) contracts, which is also not surprising for two reasons: some contracts must move up if (0,0,0) is ranked lower and the (6,14,22) contract is a "generous" contract, and principals may think agents will respond better to it if it is ranked higher in the observable treatment as we expected (even though agents failed to respond in such a way).

5.5 Hypothesis 3

Hypothesis 3 concerns the difference of principal rankings of the contracts (12,12,12), (6,14,22) and (0,0,0) between treatments. As stated above, the average ranking of these contracts differ more between the two treatments than the incentivizing contracts. Tables 5 and 6 demonstrate the percentages of time each contract was given each rank for the Hidden and Observable treatments respectively.

The distribution of rankings with the (0,0,0) contract appear consistent with Hypothesis 3 given how often it was ranked best in the hidden treatment compared to the observable treatment. Comparing the average ranking of each principal in both treatments, principals rank the contract (0,0,0) significantly better in the hidden treatment (1-sided Mann-Whitney U p=0.044). In every round, principals ranked the (0,0,0) contract in the top 3 (Best, Second, and Third) more often with the hidden treatment than the observable treatment.[19]

[18]The highest rated contract is scored as a 1 and lowest rated contract is scored as a 5.

[19]Using a Two-sided Fisher's Exact Test, this difference is significant in three of the eight rounds. Significance occurs in rounds 1,4, and 8 with p-values of 0.0055, 0.0791, 0.0266, respectively.

Table 5: **Percentage of Each Ranking-Hidden Treatment**

	(0,16,20)	**(2,18,14)**	**(6,14,22)**	**(12,12,12)**	**(0,0,0)**
Best	18.55	12.50	23.39	15.73	29.84
Second	22.58	36.29	19.76	11.69	9.68
Third	19.76	36.69	16.12	16.94	10.48
Fourth	27.63	10.08	31.05	22.58	7.66
Worst	10.48	4.44	9.677	33.06	42.33

Table 6: **Percentage of Each Ranking-Observable Treatment**

	(0,16,20)	**(2,18,14)**	**(6,14,22)**	**(12,12,12)**	**(0,0,0)**
Best	18.42	14.80	34.21	17.43	15.13
Second	26.97	28.95	21.05	17.11	5.92
Third	23.68	34.87	18.75	12.83	9.87
Fourth	21.76	14.14	17.13	37.96	11.57
Worst	9.16	7.24	8.66	14.67	57.5

The rankings of (12,12,12) and (6,14,22) appear to be consistently higher in the observable treatment compared to the hidden treatment. In particular, by comparing the average rankings of subjects between treatments, principals rank (6,14,22) significantly better in the observable treatment (1-sided Mann-Whitney U p=0.096); however, the average principal ranking displays no significant difference with the contract (12,12,12).[20]

Given the observed difference in behavior between treatments, the data indicates that principals believed that the rankings of the contracts will affect agent behavior. Agents did not appear to respond differently to the contracts between treatments despite the difference in the observability of rankings. Even though principals ranked the (6,14,22) contract significantly better and the (0,0,0) contract significantly worse in the observable treatment, agent response to these two contracts was not significantly different between the treatments. Additionally, the ranking of the (0,0,0) contract had no discernible impact on the willingness of agents to take costly effort: the rate of choosing high effort for each of the other four contracts was actually higher in the observable treatment when the (0,0,0) was ranked above last, as compared to when it was ranked last (57.5% of rankings).

[20]In addition, (6,14,22) is ranked in each of the top 3 (Best, Second, and Third) more often in the observable treatment than the hidden treatment while the flat (12,12,12) is ranked in each of the top 2 (Best and Second) more often in the observable than hidden. (6,14,22) is ranked in the top 2 more often in every round while (12,12,12) is ranked in the top 2 more often in every round except the first in the observable treatment compared to the hidden treatment. None of the differences are significant (Two-sided Fisher's Exact Test).

5.6 Hypothesis 4

Hypothesis 4 predicted that the two contracts that incentivize high effort [(0,16,20) and (2,18,14)] would be the ones most frequently ranked highest by principals. Given the much greater rate of high effort responses of agents to these two contracts, the expectation that these contracts would be highly rated was justified by agent behavior. However, our data on principal behavior is not consistent with Hypothesis 4. Principals could have ranked the (0,0,0) highly if they were worried about agent response to the incentivizing contracts, and there is some evidence of this in the hidden treatment. However, in both treatments, the contract which elicited the most high effort, (0,16,20) and provides the principal the most profit when high effort is chosen, was ranked third; additionally, the contract (6,14,22) exceeded both incentivizing contracts in terms of average ranking in the observable treatment, despite the fact it elicited less effort than both of them.

Principals behavior is clearly not consistent with profit maximization (unless principals formed incorrect beliefs on the response of agents); in addition, given our design, it does not seem to be consistent with a concern for efficiency or fairness either. In the observable treatment, the high ranking of (6,14,22) may be because principals thought agents would respond to this particular contract better if ranked higher as we hypothesized. The incentivizing contracts (2,18,14) and (0,16,20) have noisy rankings. Principals ranked (2,18,14) second or third more than seventy percent of the time in the hidden treatment and ranked it second or third more than sixty percent of the time in the observable treatment. The contract rankings for (0,16,20) was extremely noise for the top 4 rankings in both treatments. Why principals failed to recognize (0,16,20) as a better contract than (6,14,22) is unclear. They may have thought agents would react to it negatively because it is the only contract besides (0,0,0) with a payoff of zero for one of the outcomes. Thus, principals either formed incorrect beliefs on how agents would respond to the contract (0,16,20), or, given the results of the Friedman test for the first five rounds of the hidden treatment (which indicated the first five rounds were quite noisy), they failed to use adequate foresight.

5.7 Trends in the Data

Agents behavior conforming more to theory than principal behavior is unsurprising given the greater difficulty of the role of principal. Agents simply compare their own payoffs for each contract while principals must backward induct by performing the same calculation as the agent and then calculate how each contract affects the principal's own payoff. Although principals failed to act in accordance with either standard economic theory or our hypotheses, they improved over time and responded to the behavior of the agents in the experiment.

Looking over the eight periods of the games in both treatments, agent behavior has no clear trend while principals, for the most part, learned to choose contracts that earn them more over time. Appendix A provides a graph of the average principal contract rankings for

each treatment by round. Principals gave (2,18,14), (0,16,20), and (0,0,0) better rankings over time as well as giving (12,12,12) and (6,14,22) worse rankings over time. The change in rankings has a significant trend for worse rankings over time for the (12,12,12) in both treatments[21] and (6,14,22) for the observable treatment;[22] additionally, (2,18,14) has a significant trend for the better in the hidden treatment.[23]

Though significance of many of the trends is not unexpected, the lack of significance of some of the others is surprising, especially since examining the principals' decisions as agents in practice rounds showed that they behaved quite similarly to agents. Unsurprisingly, given that (12,12,12) on average provides principals with the least profit, the average ranking of the contract declined as rounds progressed. Also, the significant drop in the ranking of the (6,14,22) contract in the observable treatment likely occurred since principals may have offered it too much early in the experiment believing agents may respond more favorably to all contracts with one of the more generous contracts rated highly. Since agents only selected high effort about one third of the time with this contract, principals learned to rank it worse. Unexpected changes include the significant improvement in the ranking of (2,18,14), given that (0,16,20) elicits more effort from agents and sees less of an improvement. Even though the data shows better rankings over time for (0,16,20) over time, the absence of a significant trend for (0,16,20) is shocking. As agents in the practice rounds, principals responded with the most high effort to (0,16,20) in both treatments, yet on average it never ranked first, not even in the final round of both treatments. In fact, the principals behaved qualitatively the same in the practice rounds in the role of agents to agent behavior in the paid rounds in both of the treatments, particularly in the hidden treatment.[24] In the observable practice round, the high effort response rate to (6,14,22) was higher than the paid rounds but still lower than the rates of (0,16,20) and (2,18,14).

6 Conclusion

Our experiment tested the decision-making of both principals and agents, and agents largely behaved as predicted. Because the contracts with a monetary incentive to select higher effort far exceeded the other contracts in encouraging high effort from the agents, our results show that monetary incentives are still the biggest determinant in agent decision-making. This study is also the first to our knowledge to provide conclusive evidence that agents prefer monotonic contracts over non-monotonic contracts that have the same monetary incentives

[21] Kendall's Tau Rank Correlation Test, $\tau = -0.593$ and $\tau = -0.794$ with p-values of 0.0595 and 0.0113 for the observable and hidden treatment, respectively

[22] Kendall's Tau Rank Correlation Test, $\tau = -0.564$, p=0.0842

[23] Kendall's Tau Rank Correlation Test, $\tau = 0.667$, p=0.0327

[24] In the hidden treatment, the principals, in the role of agents of the practice rounds, responded to (2,18,14), (6,14,22), (12,12,12), (0,0,0), and (0,16,20) with high effort rates of 0.758, 0.403, 0.145, 0.130, and 0.806. In the open treatment, the principals, in the role of agents of the practice rounds, responded to (2,18,14), (6,14,22), (12,12,12), (0,0,0), and (0,16,20) with high effort rates of 0.620, 0.586, 0.246, 0.311, and 0.705.

and expected payoff; however, agents respond better to the non-monotonic contract than the non-incentivizing ones. Principals may be better offering a non-monotonic contract than an non-incentivizing or expensive monotonic contract when a monotonic contract with the same underlying incentives is not feasible.

Though monetary incentives primarily motivated agents, generosity also affected agent choice while efficiency played little if any role in agent decision-making. Roughly a third of agents did respond reciprocally to the generous, theoretically non-incentivizing contract. Unexpectedly, the flat contract that exactly compensated agents for high effort performed no better at incentivizing agents than the null contract, and the rankings of the contracts hardly affected the decision-making of the agents.

Unlike agents, principals select contracts unpredictably although some improvement occurs over time. The decisions of the principals differed greatly from both the theoretically optimal set of rankings and the optimal set of rankings given the actual agent responses in the experiment. The incentivizing contracts were theoretically the best options, and when agents responded poorly, the null contract was optimal. Principals instead consistently ranked the generous, non-incentivizing contract and the flat, effort-compensating contract too highly in both treatments, despite the fact that agents seldom rewarded these contracts with high effort. However, principals appeared to learn to better select rankings over time, especially with the poor performing flat contract.

Differences in principal rankings between treatments, particularly in regard to the ranking of the null contract, suggest that principals believed that intentions in the rankings of contracts matters. The mistakes on the part of principals were amplified in the observable treatment, as it appears principals felt pressured to rank the generous, non-incentivizing contract and the flat contract even higher and the null contract even lower than in the hidden treatment. That shift in rankings exacerbates the poor performance the principals in the observable treatment.

The lack of foresight displayed by principals in this experiment lead to both lower earnings for themselves and less efficient outcomes. Our results suggest that institutions may want to focus more on management setting the right incentives instead of reprimanding employees for following perverse incentives. Even though principals chose qualitatively in the same manner in the role of agents in the practice rounds, they failed to transition that knowledge to contract selection. Whether such poor decision-making would occur in a similar experiment with "real-world" managers or those with more training, such as MBA students, in the role of the principal is an open question. Regardless, the experimental results suggest students struggle with one round of backward induction even after playing in the role of agent in the practice rounds; thus, formal instruction on the process of contracting instead of solely experience may be required to design and select efficient contracts. Given agents generally responded to monetary incentives as predicted, the avenue for improving contracting would

require helping principals understand the impact of each contract on the agent. Because contracting occurs in every part of the economy and improving it could create enormous value, understanding how and why principals and agents behave in a given environment will always be of paramount importance.

Acknowledgments: We would like to thank Collin Weigel for helping us with the SoPHIE software, Missouri State University for providing us the funding, and participants of the 2018 ESA North American meetings for feedback and suggestions.

References

[1] Berg, Joyce E., Lane A. Daley, John W. Dickhaut, and John O'Brien. "Moral hazard and risk sharing: experimental evidence." research in Experimental Economics 5, no. 1992 (1992): 1-34.

[2] Brosig, Jeannette, Christian Lukas, and Thomas Riechmann. "The monotonicity puzzle: an experimental investigation of incentive structures." BuR-Business Research 3.1 (2010): 8-35.

[3] Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. Experimental Economics, 14(3), 375-398.

[4] Charness, G., & Levine, D. I. (2007). Intention and stochastic outcomes: An experimental study. The Economic Journal, 117(522), 1051-1072.

[5] Epstein, Seth. "Testing principal-agent theory." research in Experimental Economics 5 (1992): 35-60.

[6] Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. American economic review, 94(4), 857-869.

[7] Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. Games and Economic Behavior, 62(1), 287-303

[8] Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. The quarterly journal of economics, 108(2), 437-459.

[9] Fehr, E., Kirchsteiger, G., & Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. European Economic Review, 42(1), 1-34.

[10] Fehr, E., Naef, M., & Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. American Economic Review, 96(5), 1912-1917.

[11] Flannery, T., & Roberts, S. (2018). The use of non-monotonic contracts in a single period game: An experimental investigation. Journal of behavioral and experimental economics, 77, 177-185.

[12] Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. Econometrica, 74(5), 1365-1384.

[13] Hendriks, Achim: SoPHIE - Software Platform for Human Interaction Experiments. University of Osnabrueck, Working Paper, 2012

[14] Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. Journal of Economic Psychology, 32(5), 865-889.

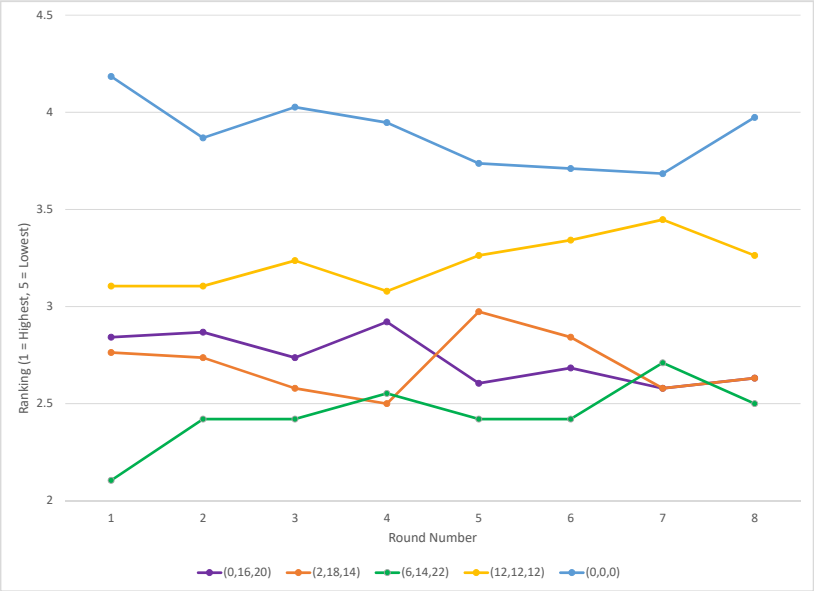
[15] Keser, Claudia, and Marc Willinger. "Principals' principles when agents' actions are hidden." International Journal of Industrial Organization 18.1 (2000): 163-185.

[16] Lukas, Christian. "Get more for less? Experimental evidence on repeated decision making under non-monotone incentives." Zeitschrift für Betriebswirtschaft 77.4 (2007): 351-379.

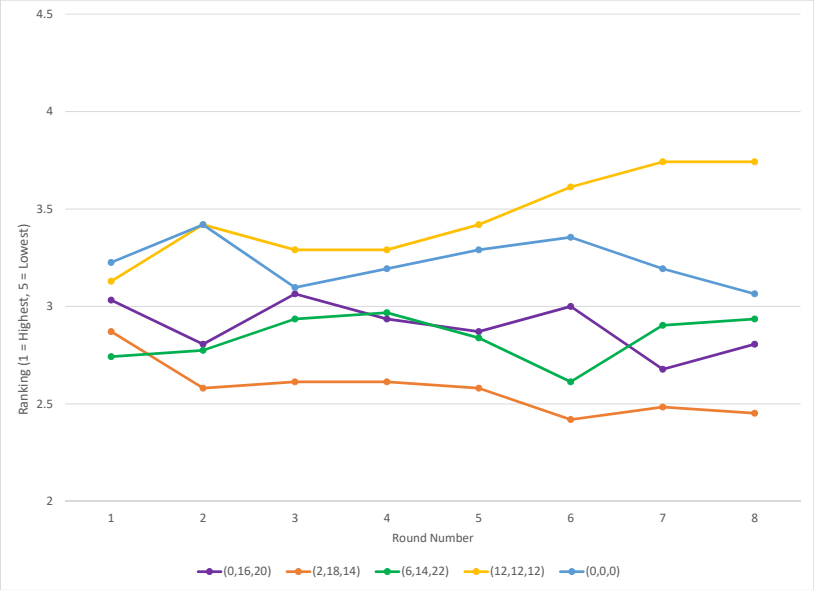
[17] Thaler, R. H., and Ganser, L. J. (2015). "Misbehaving: The making of behavioral economics." New York, NY: WW Norton: 190.

Appendix A

Average Contract Rankings by Principals: Observable Treatment



Average Contract Rankings by Principals: Hidden Treatment



Appendix B

The following instructions were given to subjects in the sessions with rankings observed by the agents.

Instructions

Table 1: Probabilities

Outcomes / **Options**	***LOW*** **Payment = 12**	***MEDIUM*** **Payment = 36**	***HIGH*** **Payment = 48**
A Cost = 0	3/4	0	1/4
B Cost = 12	0	3/4	1/4

GAME STRUCTURE

1. Player A ranks five contracts (listed below) that divide the revenue from each outcome. Each contract offers Player B a payment for three possible outcomes, LOW, MEDIUM, and HIGH. The contract Player A ranks first occurs 40% of the time, second 30% of the time, third 20% of the time, fourth 10% of the time, and fifth 0% of the time. Below are the five contracts that Player A ranks. (All contracts below listed in ECUs where 10 ECUs=$1)
   a) 2 for LOW; 18 for MEDIUM; 14 for HIGH
   b) 6 for LOW; 14 for MEDIUM; 22 for HIGH
   c) 12 for LOW; 12 for MEDIUM; 12 for HIGH
   d) 0 for LOW; 0 for MEDIUM; 0 for HIGH
   e) 0 for LOW; 16 for MEDIUM; 20 for HIGH

2. Player B chooses either "Option A" or "Option B" for each possible contract that occurs with at least a 10% chance after observing the ranking given by player A. Player B pays nothing to choose "Option A" and pays 12 ECUs from the endowment of 20 ECUs to choose "Option B." The decision of Player B influences the size of the revenue by affecting the probability of each outcome. The probability of an outcome given an action can be found in Table 1 above by looking at the intersection of an outcome and the corresponding action. For example, the MEDIUM outcome occurs three out of four times if Player B chooses "Option B".

3. An outcome is randomly generated with the probabilities in the Table 1 according to the action Player B selected (see Step 2 above) while the contract is randomly generated with the probabilities determined by Player A's ranking. Player A earns 48 ECUs for the HIGH outcome, 36 ECUs for the MEDIUM outcome, and 12 ECUs for the LOW outcome (payments **for Player A** listed in Table 1 above in the top row below the

outcomes). From these earnings, Player A pays Player B the payment offered for the given outcome according to the generated contract (see Step 1). Player B receives the payment from Player A for the given outcome in addition to the endowment of 20 ECUs minus the cost of the selected action (nothing for "Option A" and 12 ECUs for "Option B").

The game will be played twelve times, four for practice (unpaid) and eight paid. In the practice rounds, participants alternate roles between Player A and Player B. In each paid round, Player A and Player B remain in the same role but are randomly re-matched with another Player B and Player A respectively. Player identities remain anonymous throughout. At the end of the experiment, each participant receives $1 for every 10 ECUs earned in addition to the $5 show-up fee. A survey will be given at the end of the experiment as well.

The following instructions were given to subjects in the sessions with rankings hidden from the agents.

Table 1: Probabilities

Outcomes / **Options**	***LOW*** **Payment = 12**	***MEDIUM*** **Payment = 36**	***HIGH*** **Payment = 48**
A Cost = 0	3/4	0	1/4
B Cost = 12	0	3/4	1/4

GAME STRUCTURE

1. Player A ranks five contracts (listed below) that divide the revenue from each outcome. Each contract offers Player B a payment for three possible outcomes, LOW, MEDIUM, and HIGH. The contract Player A ranks first occurs 40% of the time, second 30% of the time, third 20% of the time, fourth 10% of the time, and fifth 0% of the time. Below are the five contracts that Player A ranks. (All contracts below listed in ECUs where 10 ECUs=$1)
   a) 2 for LOW; 18 for MEDIUM; 14 for HIGH
   b) 6 for LOW; 14 for MEDIUM; 22 for HIGH
   c) 12 for LOW; 12 for MEDIUM; 12 for HIGH
   d) 0 for LOW; 0 for MEDIUM; 0 for HIGH
   e) 0 for LOW; 16 for MEDIUM; 20 for HIGH

2. Player B chooses either "Option A" or "Option B" for each possible contract without ever learning the rankings chosen by Player A. Player B pays nothing to choose "Option A" and pays 12 ECUs from the endowment of 20 ECUs to choose "Option B." The decision of Player B influences the size of the revenue by affecting the probability of each outcome. The probability of an outcome given an action can be found in Table 1 above by looking at the intersection of an outcome and the corresponding action. For example, the MEDIUM outcome occurs three out of four times if Player B chooses "Option B".

3. An outcome is randomly generated with the probabilities in the Table 1 according to the action Player B selected (see Step 2 above) while the contract is randomly generated with the probabilities determined by Player A's ranking. Player A earns 48 ECUs for the HIGH outcome, 36 ECUs for the MEDIUM outcome, and 12 ECUs for the LOW outcome (payments **for Player A** listed in Table 1 above in the top row below the outcomes). From these earnings, Player A pays Player B the payment offered for the given outcome according to the generated contract (see Step 1). Player B receives the payment from Player A for the given outcome in addition to the endowment of 20 ECUs minus the cost of the selected action (nothing for "Option A" and 12 ECUs for "Option B").

The game will be played twelve times, four for practice (unpaid) and eight paid. In the practice rounds, participants alternate roles between Player A and Player B. In each paid round, Player A and Player B remain in the same role but are randomly re-matched with another Player B and Player A respectively. Player identities remain anonymous throughout. At the end of the experiment, each participant receives $1 for every 10 ECUs earned in addition to the $5 show-up fee. A survey will be given at the end of the experiment as well.