Thesis submitted to the University of East Anglia for the Degree of Doctor of Philosophy

Experimental Essays on Incentive Contracts and Obedience

Alexandros Karakostas

27th of September 2012

Acknowledgements

First and foremost, I want to thank my main supervisor Prof. Daniel Zizzo, for his invaluable advice, and continuous support throughout my PhD. I would also like to thank my second supervisor Dr. Anders Poulsen for his insightful comments and advice. I am grateful to my friend and fellow PhD student Fabio Galeotti, for his encouragement, comments and support. A special thanks needs to go to Dr. Axel Sonntag and Dr. Kei Tsutsui for their help and advice and to Melanie Parravano and Ailko Van Der Veen for assisting me running numerous sessions. I would also like to thank all my fellow PhD students, for all the interesting and amusing discussions we have had. Lastly, I owe my thanks to my family and friends for their encouragement, patience and understanding.

The financial support from the School of Economics and the CBESS is gratefully acknowledged.

Alexandros Karakostas

University of East Anglia

September 2012

Abstract

This PhD thesis consists of three essays in the field of experimental economics. The first chapter deals with the choice between three different employment contracts from a principal and the implications this choice has on the agents' behaviour. The second chapter investigates experimentally the trade-off between risk and incentives as this is described by the incentive intensity principle of Holmstrom and Milgrom (1987). Lastly, the third chapter investigates whether agents have a drive to obey as a result of social image utility towards an authority.

CONTENTS

Introduction

This PhD thesis consists of three essays on different topics within experimental economics. Although each chapter deals with a distinctively different research question, there is a broadly defined unifying theme: specifically the relationship between principals and agents.. Agency theory has been traditionally engaged with problems of hidden action and or hidden information within organizations and how to provide incentives in order to align the interests between principals and agents. Recent developments in experimental economics have been emphasising the importance of taking into account how social preferences may be affected adversely by the use of monetary incentives (e.g. Gneezy and Rustichini, 2000) as well as how incomplete contracts may lead to suboptimal outcomes not explained by standard principal agent models (e.g. Fehr et al. 1993). As a result new theoretical models have been developed which have tried to incorporate the implications of concerns for equity (Fehr and Schmidt, 1999; Englmeier and Wambach, 2010), reciprocity (Bolton and Ockenfels, 2000), and social image (Sliwka, 2007; Ellingsen and Johannesson, 2008) into standard principal-agent models.

In the first chapter of my thesis I examine how both intrinsic and extrinsic incentives can complement each other within a standard principal-agent problem of hidden action under complete information. In particular, the principal needs to choose between two different employment contracts to offer to an agent; one which is based on intrinsic and one on extrinsic motivation. Following Fehr et al. (2007), who compared a bonus contract with a monitoring contract, I compare a bonus contract with a revenue-sharing contract. The findings are strikingly different, with the revenue-sharing contract being the most preferred and most efficient contract while is considerably fair as well. In particular, the findings suggest that the earnings of both principals and agents are on average larger under the incentive contract than when the bonus contract is used but while the profits of the agents are on average the same the principal earns considerably more. In other words we observe that the principals are ripping the benefits of choosing the incentive contract as their contract choice. Nevertheless, the differences on how the surplus is shared between principals' and agents' are relatively small as, on average, the principal earns under the bonus contract the 48% of the net profits under the bonus contract and 53% under the incentive contract. Our results are in line with previous findings which suggested that social preferences matter (Fehr et al. 1998, Anderhub et al. 2002) but in contrast to findings which suggest that extrinsic motivation may crowd out intrinsic motivation (Fehr and Gachter, 2002; Fehr et al. 2007).

Whereas in the first chapter I focused on the principal-agent problem under no risk, in the second chapter I was interested on testing experimentally the theoretical predictions of the Incentive Intensity Principle of Holmstrom and Milgrom (1987). The evidence from previous studies have led to mixed results regarding the relationship between intensive intensity and risk (Prendergast, 1999). That led some scholars to describe this relationship as 'tenuous' and to attempt to provide alternative justifications of why the expected relationship is not observed in the studies (i.e. Prendergast, 2002). To our best knowledge, this is the first study which has attempted to test experimentally the trade-off between risk and incentives. Testing this relationship in a lab provides enhanced control which in turn allows obtaining more and better information on of the parameters of the model. While previous studies have found mixed evidence regarding the relationship between risk and incentive intensity, my findings are in line with the negative relationship expected by the model. In addition, I find no relation between the variance in the performance measure and the effort choice of the agent as well as a strong positive relation between the effort choice of the agents and the piece rate offered by the principal, which both are in line with the model's prediction. However, as in the literature in social preferences and gift exchange games I found that the agents responded positively to higher fixed wages, suggesting reciprocal behaviour.

While social preferences have been usefully modeled within the principal-agent framework (Englmeier and Wambach, 2010), consideration has not been given to how authority *per se* may help induce compliance. There has been considerable attention in economics to conformism and social norms, by which subjects tend to do what a number of others do (e.g. Asch, 1955; Jones, 1984; Lopez-Perez, 2008; Zafar, 2011), and there is of course a significant empirical literature on peer effects (e.g., Case and Katz, 1991; Kawaguchi, 2004; Powell et al., 2005; Lundborg, 2006) and on social image and pro-social behavior (e.g., Glazer and Konrad, 1996; Benabou and Tirole, 2006; Andreoni and Bernheim, 2009; Ariely et al., 2009), but the focus of this research has been on what may be labeled as *horizontal* social pressure, i.e. pressure by peers.

In the last chapter of this thesis I focus on whether the agent has a drive to obey as a result of social image utility towards the authority. More specifically, I examine whether *vertical*, i.e. hierarchical, social pressure induces conformism, even when there is no financial reason for obeying, and even when the domain for the action to be undertaken by the agent is anti-social. The results suggest that obedience is a powerful motivating mechanism. More specifically I find that when a constant pressure to obey is applied to engage in destruction, the destruction rate is over 40%. As many as, six subjects out of ten are willing to destroy when more pressure is provided at specific intervals in time, with no need for an explicit reason or for the potential for reciprocal aggression. As a result the findings of this chapter are important for future developments in principal-agent modeling and more generally in thinking about incentives and delegation in organizations as, even in the lack of economic incentives, individuals may tend at least to some degree to obey orders, and this is exploited by economic organizations big and small as a management tool.

Chapter 1: Efficiency and Fairness in Revenue Sharing Contracts[1]

1. Introduction

This article explores how both intrinsic and extrinsic incentives can complement each other in principal-agent relationships. The importance of monetary incentives has been repeatedly emphasised within contract theory while economists as Lazear have claimed that “Incentives are the essence of economics” (1986, p2). Most of the work in principal-agent theory has focused on how to design contracts which monetarily incentivise agents to act according to their principals’ expectations. In the meanwhile, a growing literature from experimental economics and psychology has highlighted the important role of intrinsic motivation on decision making in general, but also in the more specific context of agency problems (e.g. reciprocity and fairness concerns). This behavioural research has shown that the use of monetary incentives may undermine intrinsic motivation (Gneezy & Rustichini, 2000) or could have adverse effects in the long run (Benabou & Tirole, 2003). Whereas some authors argue for non-monetary means to overcome crucial issues embedded in principal-agent settings such as selecting agents by their preferences (Prendergast, 2008), other researchers pointed out that monetary incentives might work, however their effectiveness depends on the degree of the agents’ intrinsic motivation (Boly, 2010) or the size of the monetary incentives (James, 2005). In this paper we integrate the above perspectives by showing that intrinsic and extrinsic motivation may not crowd out each other but instead act as complements, leading to Pareto optimal allocations (c.f. Murdock, 2002).

Recently, Fehr et al. (2007) showed that most principals, when offered the choice between an enforceable monitoring contract (MC) and a non-enforceable bonus contract (BC), preferred the bonus contract (roughly 90%). In addition, they found that the effort exerted by the agents and the average payoff for both the

[1] This chapter is based on a paper co-authored with my supervisor Prof. Daniel Zizzo and Dr. Axel Sonntag.

principals and the agents were higher in bonus contract than in monitoring contract settings. These results are the opposite of what economic theory predicts under a narrow self-interest assumption. In particular, payoff maximizing principals should choose the monitoring contract as its anti-shirking punishment mechanism should result in higher effort and payoff levels than in a bonus contracts setting (where effort is expected to be zero). Fehr et al. (2007) argued that the bonus contract was preferred to the monitoring contract due to social preferences. However, other reasons might have affected the principals' choices as well. Regarding Fehr & Gächter (2002), selecting of a contract that contains the possibility of fining could be perceived as a hostile act itself and might send the agent a signal of distrust which in turn could have increased the likelihood to shirk (Bacharach et al., 2007). Hence, the monetary incentive of a 33% chance of detecting shirking behaviour (as it was the case in Fehr et al. (2007)) was probably not enough to outweigh the disadvantageous effects of detrimental behavioural signals. In contrast to the results of Fehr et al. that cannot be explained by narrow profit maximizing behaviour, Anderhub et al. (2002) who used a similar agency setting, found that principals, when using highly flexible revenue sharing contracts, "clearly recognize the agency problem and react accordingly" (Anderhub et al., 2002, p.24). In other words, principals behaved very much in line with the profit maximizing predictions. However, interestingly, they also "take fair sharing into account" (Anderhub et al. 2002, p.24).

In order to consolidate previous findings that a contract with voluntary bonus payments dominated a contract that offered enforceable monetary incentives with evidence that positive monetary incentives could achieve efficient yet fair outcomes, we combine Fehr et al.'s (2007) bonus contract with a less flexible version of Anderhub et al.'s (2002) revenue sharing contract[2] and add a trust contract as a third option. In our revenue sharing contract a principal defines a non-negative fixed wage and additionally offers the agent a share of the total (gross) revenue. In contrast to the MC, this contract does not involve any payoff-relevant probabilities and therefore does not induce additional risk. As the principals also lack an instrument to specify fines for shirking agents, they do not need to worry

[2] We do not allow negative fixed wages.

about sending any negative signals of distrust or hostility. Most importantly, the revenue sharing contract (henceforth Incentive Contract or IC) allows social preferences to be expressed by offering a generous share of the total revenue to the agent. Hence, a principal can express social preferences in both the BC and IC if she chooses to do so.

Considering the above, there are good reasons to expect bonus and incentive contracts to appear interesting to the principals. Finding out which contract finally is preferred over the other is the main aim of this article. Additionally, it is also of interest to examine which contract on average will generate the highest total surplus and how that surplus will be divided between the principal and the agent. Furthermore, there are two additional dimensions which are distinctively different in our design from that of Fehr et al., (2007) and Anderhub et al., (2002). Firstly, in our experiment the principal can choose between three possible contracts. Those are the trust contract (TC), the bonus contract (BC), and the incentive contract (IC). In the TC the principal offers a fixed wage to the agent first, who in response exerts a specific effort level. The structure of the trust contract therefore can be interpreted as a gift exchange game (c.f. Fehr et al., 1998). Fehr et al. (2007) pointed out that the trust contract is just a special case of a bonus contract (as one could choose the BC and set the bonus to zero) and Anderhub et al. (2002) similarly argued that TC is also just a special case on an incentive contract (as one could only pay a fixed wage and set the revenue share to zero). Yet, there may be motives for a principal who chooses the BC or IC to actually use the bonus or the revenue share as it is an available option. In order to investigate the usage of such opportunity we made the TC an explicit option.

Secondly, and perhaps more importantly, we controlled for confound effects between the contract choices and determining the parameters within the chosen contract. When a principal has the option to choose among the three contracts the choice of the contract per se may act as a signal to the agent regarding the intentions of the principal. For example, when a principal chooses the BC among the three contracts and promises a large bonus with a small fixed wage for a high effort level, this may be interpreted by the agent as attempting to fool him (the agent) into exerting a high effort in order to free ride on him afterwards. This interpretation is conditional to the fact that the principal could have chosen the IC in which it could have been ensured that the agent is paid for his effort. However, when the BC is exogenously given, the intentions of the principal are less clear. More generally, when there is a range of contracts the choice of contract may create a signal which consequently will affect both the parameters chosen within the contract and the behaviour of the agent. In order to control for this potential confound games with exogenously set trust, bonus, and incentive contracts have been added to the experimental design.

The remainder of this article is structured as follows: Section 2 presents the theoretical predictions. Section 3 describes the experimental design. Section 4 presents and discusses the results, and section 5 concludes. The experimental instructions as well as proofs for the theoretical prediction are provided in the appendix.

2. The Principal Agent Problem and Contract Design

In the chosen setting a principal hires an agent to carry out production. The revenue of production depends on the agent's effort level e such that $R(e) = 150 * e$. For providing effort, the agent bears a cost of $C(e) = e + e^2$ with $e \in \{0,1,\ldots,19,20\}$. Neither on revenue nor on effort can be contracted upon. In our environment exist three types of contracts.

Trust contract (TC)

In a trust contract the principal offers the agent an unconditional fixed wage F and suggests the agent to provide an effort level e^s. However, if the agent accepts this offer, the suggested effort level e^s cannot be enforced by the principal. Consequently, the principal's monetary payoff resulting from a trust contract $TC(F, e^s)$ is defined as $\mathrm{R(e)} - \mathrm{F}$, whereas the agent earns $\mathrm{F} - \mathrm{C(e)}$.

Bonus contract (BC)

Similarly to the trust contract, the principal offers a fixed wage F and suggests an effort level e^s. However, differently to the TC, the principal also announces to pay a bonus $B \in \{0,1,\ldots,2999,3000\}$ if the agent delivers the suggested effort level. After the agent's effort choice, the principal has the opportunity to pay the agent a *voluntary* bonus in addition to the fixed wage F.

Neither the agent's effort level e nor the principal's bonus payment B are enforceable.

Incentive contract (IC)

The principal jointly decides on a fixed wage F and a share $S \in \{0.00, 0.01, \dots ,0.99,1.00\}$ that specifies how much of the totally generated revenue she *will* return to the agent. Note that in contrast to the bonus payment B, the amount $R(e) * S$ is compulsory part of the incentive contract, i.e. there is no uncertainty involved whether the principal might pay a bonus and – if at all – what amount, but the agent by choosing his effort level also *determines* his own income. The agent's payoff in our incentive contract setting is $\mathrm{F + S * R(e) - C(e)}$. Conversely, the principal's payoff is $\mathrm{R(e) * (1 - S) - F}$.

The expected payoffs for principals and agents are to be understood as on top of their initial endowment.

$$\mathrm{P^P} = \begin{cases} \mathrm{R(e) - F} & \text{for TC} \\ \mathrm{R(e) - F - B} & \text{for BC} \\ \mathrm{R(e) * (1 - S) - F} & \text{for IC} \end{cases}$$

$$\mathrm{P^A} = \begin{cases} \mathrm{F - C(e)} & \text{for TC} \\ \mathrm{F + B - C(e)} & \text{for BC} \\ \mathrm{F + S * R(e) - C(e)} & \text{for IC} \end{cases}$$

Under the assumption that both, principals and agents behave as selfishly payoff maximizers, the game theoretic solution predicts that principals should never choose the trust contract nor the bonus contract. In both contracts, TC and BC, selfish agent has no incentive to provide a higher effort level than zero. As the principal anticipates this incentive structure, she should never offer such a contract with a fixed wage greater than zero. Hence, following standard economic theory, neither TC nor BC should result in Pareto improvements relative to the status quo. The incentive contract, however, due to its obliging structure potentially provides sufficient monetary incentives to achieve a net welfare increase (Pareto improvement). The word 'potentially' points to the fact that not all feasible incentive contracts are incentive compatible. In order to make an IC incentive compatible, the return share of the total revenue needs to exceed a certain value.

Therefore, selfishly payoff maximizing principals should choose the incentive contract, suggest an effort level of 20 and offer a revenue share of 0.27[3].

3. Experimental Design and Hypotheses

At the beginning of the experiment all subjects were randomly selected to become principals or agents and were paired up to groups consisting of one principal and one agent each (absolute stranger fixed matching). It was common knowledge that this role would not change until the end of the experiment. The instructions were split-up into several parts and were provided to the subjects in a piecewise manner at the beginning of every stage. The experiment consisted of several different games the order of which was subject to experimentally controlled variation. The following paragraphs describe each of the games. Figure 1 presents a structural overview of the full experimental design.

Figure 1: Sequential Structure of the Experiment

Sessions 1-6

Round 1	Round 2-7	Round 8	Round 9	Round 10
TBI	TBI-r	ex.TC	ex.BC	ex.IC

Sessions 7-12

Round 1	Round 2	Round 3	Round 4	Round 5-10
ex.TC	ex.BC	ex.IC	TBI	TBI-r

Note: In the games ex.TC, ex.BC, and ex.IC the contracts could not be chosen but were set exogenously to be a Trust, Bonus, and Incentive Contract, respectively. TBI and TBI-r represent a one shot and repeated contract choice settings, respectively.

[3] The derivation of these contract specifications is presented in the appendix.

The exogenous Trust Contract game (ex.TC)

In the first stage of the TC game the principal chooses the size of the fixed wage she wants to offer and suggests an effort level to the agent. In the second stage the agent after being informed about the offered contract decides on an effort level. In the third stage both the agent and the principal get informed about their earnings and the round ends.

The exogenous Bonus Contact game (ex.BC)

In the first stage of the BC game the principal chooses the size of the fixed wage she wants to offer, and the size of the bonus she promises to offer if she is satisfied by the effort level that will be offered by the agent and also suggests an effort level to the agent. In the second stage the agent after being informed about the offered contract decides an effort level. In the third stage the principal after being informed about the agent's effort level decides if she wants to offer a bonus and of what size. In the fourth stage both the agent and the principal get informed about their earnings and the round ends.

The exogenous Incentive Contract game (ex.IC)

In the first stage the principal decides the values of the fixed wage, the share on the total revenue that will be given to the agent, and suggests an effort level. In the second stage the agent after being informed about the offered contract decides an effort level. In the third stage both the agent and the principal get informed about their earnings and the round finishes

The one shot Trust-Bonus-Incentive (TBI) and repeated Trust-Bonus-Incentive (TBI-r) games

In the first stage the principal chooses between the 3 possible contracts. In the second stage the principal decides the values for the parameters of the chosen contract and suggests an effort level to the agent. In the third stage the agent after being informed about the offered contract decides on an effort level. In the fourth stage both the agent and the principal get informed about their earnings and the round finishes. In the repeated version of the TBI game the aforementioned process is repeated five times (resulting in six TBIr games in total). After the subjects read the instructions they had to answer four multiple choice questions. These were programmed in z-Tree (Fischbacher, 2007) and explanations of the right or wrong answer were provided by the program. Afterwards, three practice games identical to the ex.TC, ex.BC, and ex.IC were played to help the subjects familiarising with the experiment and the z-Tree environment. The participants could ask questions at the end of each practice round. The three trial rounds were followed by ten actual payoff-relevant rounds consisting of one one-shot game of ex.TC, ex.BC, ex.IC, and TBI, amended by 6 repeated TBI games (TBI-r). The order of the ten actual rounds was counterbalanced over sessions (see Figure 1) to control for order effects. As it has been shown above, under the assumption of profit maximizing behaviour, neither the trust contract nor the bonus contract would result in payoffs other than the initial endowments. In contrast, the incentive contact offers the opportunity to achieve much higher profits both for the principals and the agents. Therefore, it is expected that:

Hypothesis 1: Principals prefer the incentive contract over the bonus or trust contract.

Hypothesis 2: Agents provide more effort in incentive contract than in bonus contract or trust contract settings.

Assuming profit maximizing behaviour of both principals and agents should result in a 60:40 split of the profits in favour of the principals; a distribution clearly different from an equal split. However, as it is well known from experimental literature that individuals do actually show consider fairness as well, we might be finding a more equal distribution than the profit maximizing prediction.

Hypothesis 3: The overall profits are distributed in a ratio of 60:40 between principals and agents to the disadvantage of the latter.

Profit maximizing principals should, whenever allowed to do so, choose the incentive contract and offer an incentive compatible share of the generated revenue. However, they should not offer more than the lowest share that incentivises the desired effort level sufficiently. Therefore, it is expected that:

Hypothesis 4: Principals offer the lowest feasible incentive compatible share of 0.27.

The incentive contract is the only contract that theoretically allows a Pareto improvement compared to the initial endowments. Thus, it is expected that:

Hypothesis 5: Principals' and agents' combined surplus is higher in incentive contract than in bonus or trust contract settings.

4. Results and Discussion

The actual experiment was conducted in the laboratory of the Centre for Behavioral and Experimental Social Sciences of the University of East Anglia. In total 144 subjects participated in the experiment. On average participants received £15.46 for an effort of approximately 90 minutes. All subsequently presented non-parametric test statistics were calculated on session level means per game (ex.TC, ex.BC, ex.IC, TBI, or TBI-r)[4], unless stated otherwise. As some experimental observations are not independent from each other (which is required by the applied statistical tests), this procedure is necessary to attain unbiased test statistics.

Result 1: In line with hypothesis 1, when given the option to choose between three contracts, both in the one shot (TBI) and the repeated games (TBI-r) a great majority of subjects chose the incentive contract IC.

Evidence: Whereas on average 75% decided to choose the Incentive contract, the Bonus contract and the Trust contract were chosen much less frequently, accounting for 21% and 4% of total choices, respectively[5].

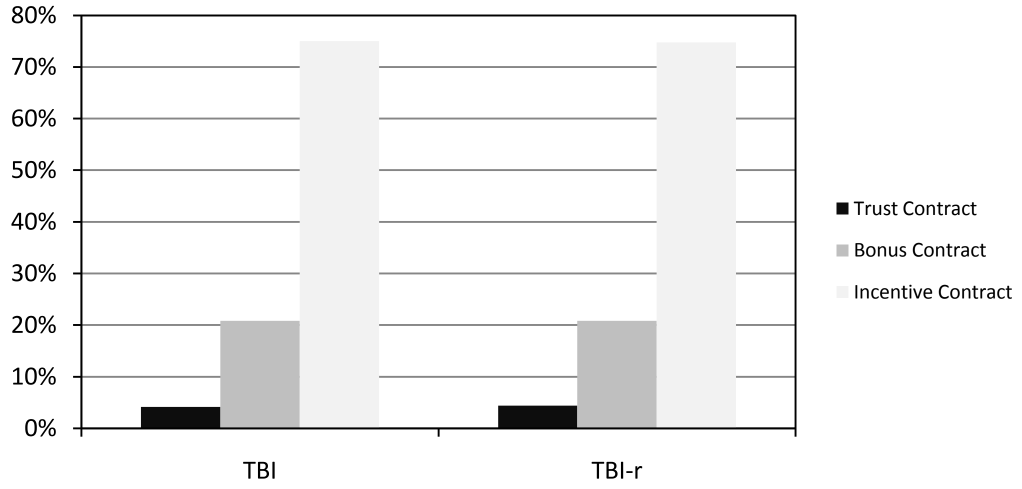
Although the relative share of Trust, Bonus and Incentive contracts seems to be quite different, interestingly, the distribution among contract choice between one shot (TBI) and repeated (TBI-r) contract choices is not different at all (see Figure 2). Computing Wilcoxon ranksum tests for differences between TBI and TBI-r for each of the three contracts resulted in non-significant results with $p > 0.9$ for all

[4] As there is only one observation for every ex.TC, ex.BC, ex.IC, and TBI round, the means equal the individual observations. Only the data from TBI-r (as repeated six times) is affected by this procedure.

[5] These numbers do not vary between TBI and TBI-r. Corresponding non-parametric statistics that test for potential differences between TBI and TBI-r are provided subsequently.

three comparisons[6]. This result is surprising as one might have expected a higher share of bonus contracts in the repeated game (TBI-r), as reputation effects might come into play (c.f. Falk et al. 1999).

Figure 2: Percentages of chosen contracts in one shot (TBI) and repeated choice (TBI-r) settings

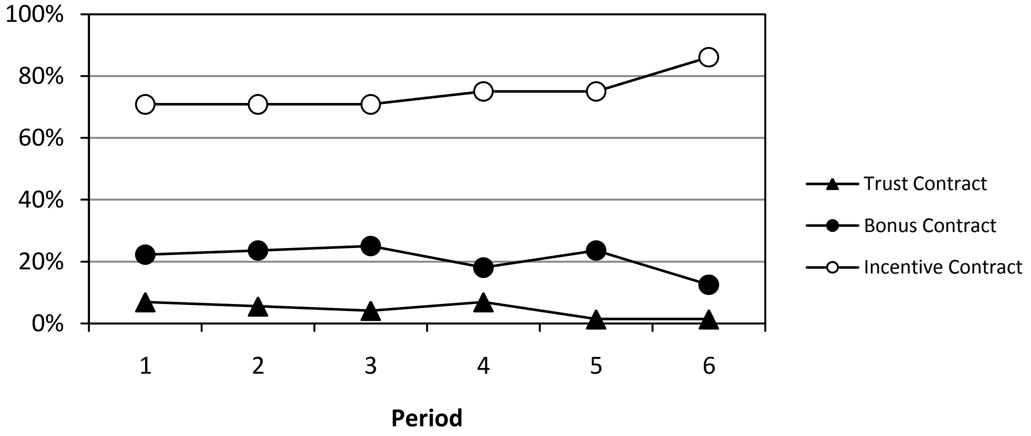


However, if building trust between principals and agents might have played a role (fixed matching in TBI-r for six rounds), it did not seem to affect the principals' contract choice behaviour. Investigating data from the one shot contract choice (TBI) for differences regarding relative contract choice propensities results in highly significant results between all possible combinations (TC<->BC: p=0.003, TC<->IC: p<0.001, BC<->IC: p<0.001). Thus, IC was significantly more often chosen than BC which itself had a significantly higher propensity to be chosen than TC. Computing Wilcoxon ranksum tests for the TBI-r data reveals similar results. In fact, IC was statistically significantly chosen more often than BC (p<0.001) and TC (p<0.001), whereas the frequency of BCs was significantly higher than the frequency of TCs (p<0.001).

[6] Note that individual observations were used to compute these Wilcoxon ranksum test statistics as using the averages of TBI-r observations would lead to wrong conclusions. This is because averaging affects rank comparisons considerably, and thus would results in a not directly comparable vector of contract choices.

Figure 3 focuses on repeated choices and illustrates the distribution of the contracts across time for the repeated choice setting (TBI-r). The graph indicates that on average the IC is chosen more often in the later rounds. Starting at 71% on the first game and gradually increasing to 86% on the sixth and last game. In order to test whether there exists a significant time trend, a simple probit model was estimated with IC choices as dependent and the experimental TBI-r period as independent variable. This estimation results in a statistically significant positive time effect (p=0.03). Analyzing the corresponding marginal effects indicates that the probability of choosing IC increases by 2.7% in each period. There was no significant time trend for choosing BC (probit regression, negative coefficient but p=0.16) nor for the share of trust contracts (probit regression, p=0.06). Principals seem to realize over time that choosing the incentive contract could increase their payoff considerably (see result two). Interestingly, although using ten repetition periods, which should increase learning effects, Fehr et al. (2007) did not find a significant time trend for the bonus contract that was dominantly chosen in their experiment. In summary, as the incentive contract was preferred over the bonus and the trust contract in all contract choice settings, result 1 clearly corroborates hypothesis 1.

Figure 3: Percentage of chosen contracts over time in repeated contract choice settings (TBI-r)



Result 2: In line with hypothesis 2, the incentive contract was the most efficient contract in terms of agents' effort and total revenue.

Evidence: As the principals' profit strictly monotonically increases with the effort applied by the agents, suggesting the maximum possible amount (20) would be the only meaningful decision.

Figure 4: Suggested effort by contract type over time in repeated contract choice settings (TBI-r)

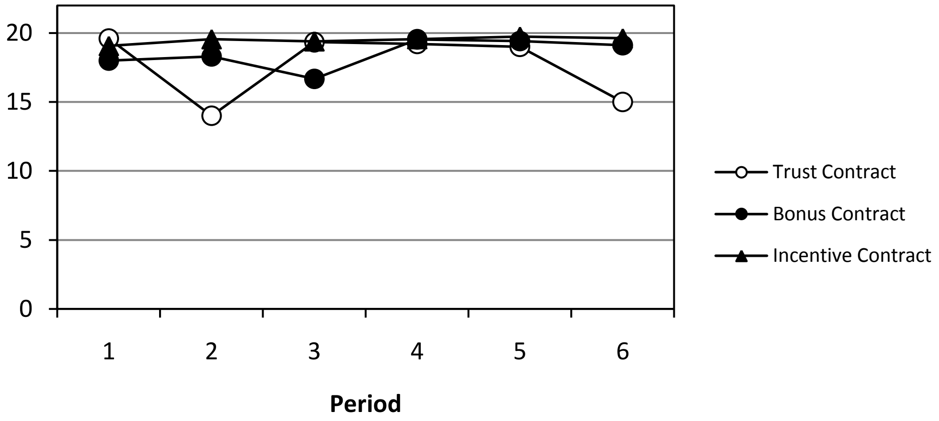
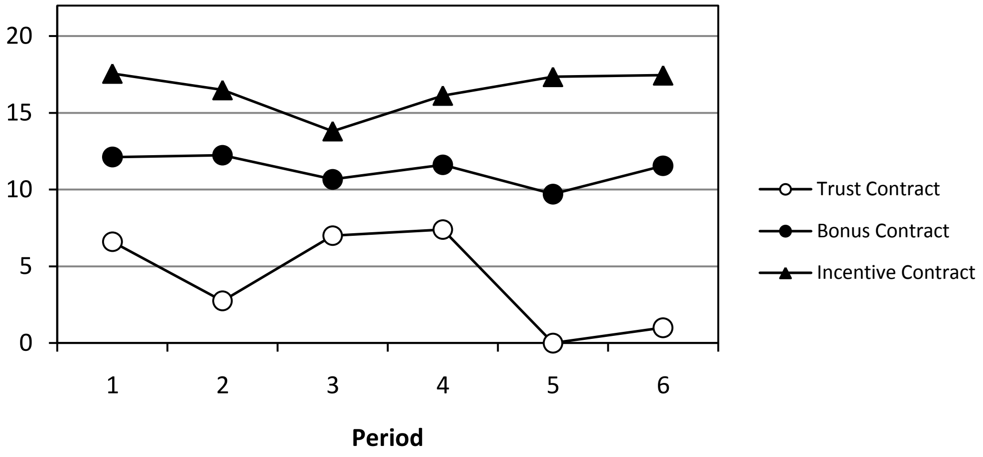


Figure 4 illustrates that most principals understood what would be most beneficial to them, resulting in very high levels of demanded effort. There is also very little variation between contract types, which is in line with the theoretical prediction.

Figure 5: Revealed effort by contract type over time in repeated contract choice settings (TBI-r)



Analyzing agents' actual effort response, however, gives a quite different picture (see figure 5). There seem to be huge differences between the three possible contracts. Due to the linear relation between the agent's effort and the resulting total revenue[7], a more detailed statistical analysis of revealed effort is deliberately omitted here in favour of more descriptive details and non-parametric tests on total revenue, subsequently. On average, the incentive contract IC generated approximately 68% more total revenue than the bonus contract in the one shot settings (TBI) and 45% more total revenue than BC in the repeated choice (TBI-r) situations. The trust contract TC, as expected, is by far the worst alternative with respect to efficiency. Table 1 reports the average total revenue generated within each chosen contract type in the TBI and TBI-r, respectively.

Table 1: Average total revenue by contract type in one shot (TBI) and a repeated choice (TBI-r) settings

	Trust Contract	**Bonus Contract**	**Incentive Contract**
TBI	200	1420	2389
TBI-r	813	1690	2476

Comparing the average total revenues of the BC and the IC contracts in the TBI and TBI-r rounds reveals that the incentive contract in the one shot (TBI) as well as in the repeated (TBI-r) game, results in higher total revenues (see Table 1). A Wilcoxon ranksum test on these differences between IC and BC revealed statistically significant results both for the TBI ($p=0.003$) and TBI-r rounds ($p<0.001$), describing IC as the more efficient contract. In addition, to test whether the average total revenues of BC and IC are statistically different between the TBI and TBI-r rounds, Wilcoxon ranksum tests were calculated on that as well. Whereas the test statistic for the IC comparison was significant with $p=0.006$, there could not be found significantly higher revenue for BC in the TBI-r as compared to TBI

[7] Revenue = 150 * Revealed Effort

(p=0.421). However, this only might be due to too few observations for BC in the TBI games. Considering the above evidence, hypothesis 2 is corroborated.

Result 3: In contrast to hypothesis 3, both in the one shot (TBI) and the repeated (TBI-r) contract choice settings, when the incentive contract was chosen, the distribution of the total surplus was fairer than the profit maximising prediction. This result is driven by the principals' choice of providing higher than efficient revenue shares.

Evidence: In the previous sections it has been shown that the incentive contract was chosen most often compared to the other alternatives BC and TC. Additionally, IC generated significantly higher revenues than the other contract options. However, the actual distribution of profits has not been discussed yet. Before reporting and analyzing empirical evidence on actual profit distributions and the implied degree of fairness, it is worth considering the profit distributions predicted by economic theory. Profit maximizing theory for trust contracts as well as for bonus contracts predicts zero effort by the agents, resulting in zero revenue and consequently no additional gains compared to the status quo (endowments). In contrast, the incentive contract encourages selfish agents to show effort and is expected to produce a net surplus compared to the status quo. In particular, if both, the principal as well as the agent would behave as if they were profit maximizers, the principal would receive 5.190ECU and the agent's payoff would be 3.390ECU (see details about the payoff-structure in the appendix).

The top part of Table 2 shows that the principal would receive 60% of the total surplus, leaving only 40% for the agent, a clearly unequal split. However, additionally Table 2 also displays the actually observed profit distributions in a one shot (TBI) and repeated (TBI-r) choice situations. The results suggest that on average the distribution of the total surplus, when the incentive contract was chosen, was fairer than what economic theory would predict in both TBI and TBI-r. In particular, the principals on average kept 53% of the total surplus in the TBI rounds, and 54% in the TBI-r, respectively. Whereas the distribution of profits in both TBI and TBI-r were significantly different from the theoretical prediction (Wilcoxon test, both p<0.001), there was no significant difference between TBI and TBI-r (Wilcoxon test, p=0.150). This result rejects hypothesis 3 (that the split would be 60:40) and is, again, rather surprising as it might be expected that in a finite repeated game reputation effects could promote social preferences (Falk et al. 1999) which might be reflected in more equal distributions of profits. Comparing our profit distribution with other empirical findings of the similarly structured settings indicates that a division of 53:47 split could be considered as a fair split (Fehr & Schmidt 1999; Güth et al. 1982). Such a share might even be regarded as surprisingly equal, taking into account that principals might rightly claim more than half, as it was them who, by their contract choice, made any gains in total revenue possible in the first place (Güth et al. 1982).

Table 2: Distribution of profits resulting from incentive contracts for the profit maximizing prediction, one shot (TBI), and repeated choice (TBI-r) settings

		Average Profit	**Relative Share**
Theoretical Prediction	**Principal**	5190	60%
	Agent	3390	40%
	Difference		21%
Results TBI	**Principal**	4269	53%
	Agent	3806	47%
	Difference		6%
Results TBI-r	**Principal**	4359	54%
	Agent	3786	46%
	Difference		7%

Result 4: In line with hypothesis 4, the great majority of offered incentive contracts were specified incentive compatible. However, additionally, principals offered significantly higher revenue shares than would be efficient.

Evidence:

Table 3 provides information on the incentive compatibility of incentive contracts observed across all sessions. In the exogenous IC round8, i.e. principals

[8] For more details see the experimental design in section 3.

only had to specify the parameters of the incentive contract, but could not choose between TC, BC, and IC, out of the 72 observed incentive contracts 53 (74%) where incentive compatible and 19 (26%) were not.

Table 3: Share of incentive compatible and incentive incompatible incentive contracts by experimental game types

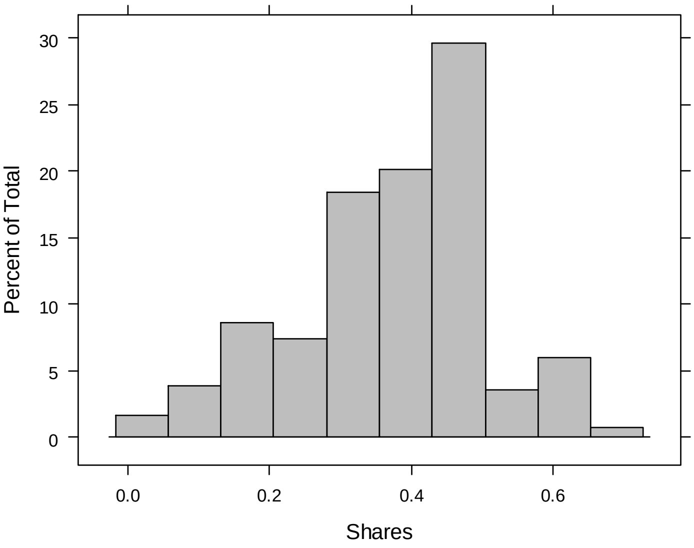
	Number of Observations	**incentive compatible**	**incentive incompatible**	**incentive compatible %**	**incentive incompatible %**
Exogenous IC	72	53	19	74%	26%
TBI	54	45	9	83%	17%
TBI-r	323	253	70	78%	22%

Note that *Exogenous IC* denotes a one shot incentive contract game with an exogenously determined contract type (IC). TBI and TBI-r represent contract choice situations in a one shot and repeated game, respectively.

In the one shot contract choice setting (TBI) 54 incentive contracts were observed of which 45 (83%) where incentive compatible and 9 (17%) were not. Similarly, in the repeated contract choice setting (TBI-r) 253 out of 323 (78%) incentive contracts where incentive compatible. Note that an IC was deemed incentive compatible if the offered share was equal or higher to 0.27. The throughout relatively high proportion of incentive compatible contract offers indicate that principals in general had a good understanding of what contract specifications they were providing. Moreover, interestingly, the principals on average offered a considerably higher share of their revenues than it would make sense if both, principals and agents were selfish profit maximizers.

Figure 6 indicates that the majority of revenue shares offered were higher than the efficient level of 0.27. In fact, the mean offered share was 0.382 (median 0.4), and was not significantly different between ex.IC, TBI, and TBI-r games. The actually offered shares were statistically significantly higher than the lowest incentive compatible share 0.27 (Sign test, p<0.001), which clearly rejects hypothesis 4.

Figure 6: Histogram of revenue shares offered by the principals of all accepted incentive contract games



Note that *all accepted incentive contract games* includes exogenously set ICs (ex.IC), one shot (TBI) and repeated contract choice situations.

Considering that in an IC setting the principals could influence the distribution of final profits by setting fixed wage and revenue share, the actually observed share offers might mean that principals actually do care about fair outcomes. Raising the revenue share over the efficient incentive compatible level 0.27 cannot be explained by selfish profit maximization. Hence, result 3 in general suggests that under an incentive contract principals may actually show concerns for fairness. Such an interpretation would be in line with the results of Anderhub, *et al.* (2002).

Result 5: In line with hypothesis 5, using incentive contracts on average led to a significantly higher total surplus than using bonus or trust contracts. However, such welfare gains were mostly absorbed by the principals.

Evidence: Table 4 reports the average profits of TC, BC and IC as observed in one shot setting (TBI). Examining the absolute surplus of agents is not significantly different between BC and IC (Wilcoxon ranksum test, p=0.343). The same is true for the difference between agent's surplus resulting from TC and BC (p=0.953), as well as, TC and IC (p=0.111). However, the principals' profits seem to vary quite considerably between different contract situations. The incentive contract led to significantly higher profits for the principals than the trust contract (p=0.004) or the bonus contract (p=0.030). Interestingly, the difference between TC and BC was not statistically significant (p=0.172). In general, the results displayed in Table 4 indicate that principals absorb a considerable share of the welfare gains obtained by choosing a more efficient contract type.

Table 4: Average and standard deviation of profits in one shot contract choice settings (TBI)

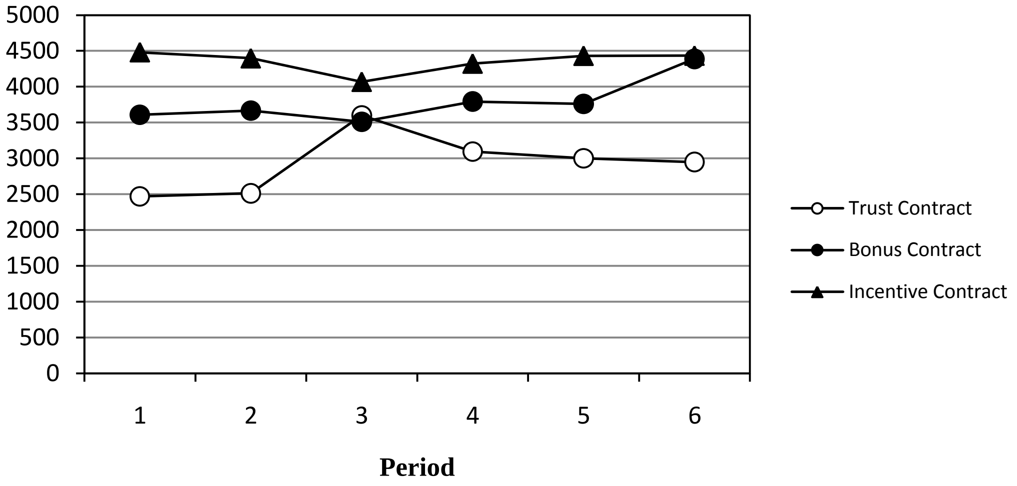
	Trust Contract		**Bonus Contract**		**Incentive Contract**	
	Av. Profit	*SD*	*Av. Profit*	*SD*	*Av. Profit*	*SD*
Principal	2700 (44%)	173	3582 (48%)	1205	4269 (53%)	577
Agent	3496 (56%)	3	3663 (52%)	542	3806 (47%)	410
total	6196 (100%)	170	7245 (100%)	1190	8075 (100%)	730

Similar results are also maintained in the repeated choice setting (TBI-r) as displayed in Table 5. There is a fairly small deviation of the distribution of the total surplus between BC and IC, and a larger difference in the TC. In addition, as before the principals who used the IC received the largest share of the total surplus, compared to BC and TC. However, noteworthy, the increase of the principals' relative profit share did not have detrimental effects on the agents' payoffs. Investigating the absolute surplus of agents is not significantly different between BC and IC (Wilcoxon ranksum test, p=0.425). The same is true for the difference between agent's surplus resulting from TC and BC (p=0.609), as well as, TC and IC (p=0.723). There is also no significant difference of agents' profits between TBI and TBI-r no matter which contract was chosen. Wilcoxon ranksum tests between TBI and TBI-r result in p-values for the TC, BC, and IC comparisons of 0.249, 0.518, and 0.157, respectively.

Table 5: Average and standard deviation in repeated contract choice settings (TBI-r)

	Trust Contract		**Bonus Contract**		**Incentive Contract**	
	Av. Profit	*SD*	*Av. Profit*	*SD*	*Av. Profit*	*SD*
Principal	2875 (42%)	1076	3733 (49%)	988	4359 (53%)	651
Agent	3849 (58%)	668	3755 (51%)	660	3786 (47%)	428
total	6725 (100%)	995	7489 (100%)	1136	8145 (100%)	768

Figure 7: Average principals' profits by contract type over time in repeated contract choice settings (TBI-r)



Thus, it can be concluded that the profit of agents did not change no matter if the game was repeated or not (TBI-r vs. TBI) and no matter which contract was chosen. Differently to the agents' profits, principals' profits varied with respect to contract choice. In the repeated choice setting (TBI-r) both, BC and IC significantly increased principals' profit compared to TC (Wilcoxon ranksum test, both p<0.001).

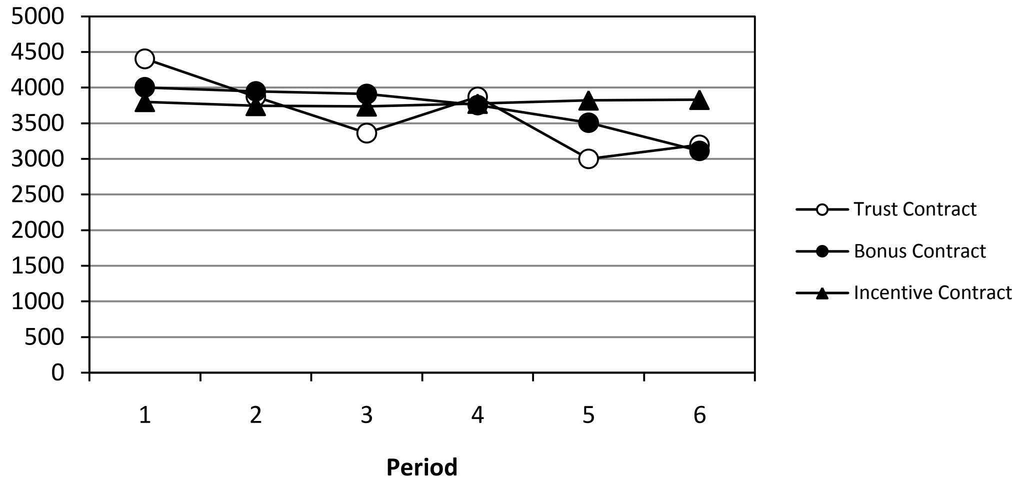
Figure 7 and Figure 8 illustrate the average profits in TBI-r of principals and agents, respectively. Whereas the there is a clearly visible spread between the principals' average profits, agents' profits do not seem to vary much with respect to different contracts. Moreover, IC significantly outperformed BC (p<0.001) with respect to principals' average profits. Similarly to the one shot choice (TBI), also in the repeated settings (TBI-r), any efficiency gains seem to be consumed by the principals only. Agents do not seem to profit from a growing cake, however, they also do not receive less. Consolidating the above evidence, the combined profit of principals and agents was significantly higher in incentive contract than in bonus and trust contract settings, which is in line with hypothesis 5.

Result 6. The bonus contract is a riskier alternative than the incentive contract as measured by variance in resulting profits.

Evidence: Considering Table 4 and Table 5 reveals that the standard deviation of the average of total profits both for principals and agents is much larger for the bonus contracts (BC) than for the incentive contracts (IC). It has been argued earlier that one of the reasons why IC could dominate BC is that it is perceived as a contract involving less risk as compared to BC.

Recall that in the incentive contract the share of profits is part of the contract which allows both the principal and the agent to know what exactly their earnings will be for each level of effort before the agent actually decides on his/her effort. In contrast, in the bonus contract setting the principal does not realise what her profits are going to be until the agent has decided an effort level, while the agent realises his profits only after the principal decides what bonus she wants to pay, after she has observed the actual effort level. Because of these inherit elements of the contract designs one would expect the variance of principals' profits to be larger in bonus contract situations. Consequently, risk averse principals could be deterred from choosing the bonus contract.

Figure 8: Average agents' profits by contract type over time in repeated contract choice settings (TBI-r)



In order to test whether the variance of the principal profits in the BC is larger than in the IC, three tests have been used. A Levene's test (1960) and two alternative formulations of Levene's test which where suggested by Brown & Forsythe (1974). Levene's test relaxes the assumption that the data are drawn from an underlying normal distribution (which would be required for an F test). The two

Table 6: Determinants of Agents' Effort

Dependent Variable:	(1)	(2)	(3)	(4)
Revealed Effort	Robust Standard Errors	Robust Standard Errors	Hierarchical Linear Model grouped by subjects and sessions	Hierarchical Linear Model grouped by subjects and sessions
Constant	-5.006 (2.937)	-5.761 (6.999)	-5.004 (2.936)	-5.760 (6.999)
Demanded Effort	0.460*** (0.122)	0.508 (0.423)	0.460*** (0.122)	0.508 (0.423)
Fixed wage	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)
BC	1.373 (2.703)	3.509 (7.577)	1.372 (2.703)	3.511 (7.577)
BC x Demanded Effort		-0.136 (0.451)		-0.136 (0.451)
BC x fixed wage	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)
BC x announced bonus	0.006*** (0.001)	0.006*** (0.001)	0.006*** (0.001)	0.006*** (0.001)
IC	7.837** (2.461)	5.732 (8.125)	7.837** (2.461)	5.730 (8.125)
IC x Demanded Effort		0.107 (0.475)		0.107 (0.475)
IC x fixed wage	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)
IC x share	9.407** (3.290)	9.307** (3.297)	9.409** (3.291)	9.309** (3.298)
IC x Incentive compatible offer	2.823** (1.059)	2.649* (1.079)	2.823** (1.059)	2.648* (1.080)

This table reports the coefficients and robust standard errors of OLS and hierarchical linear regressions for repeated choice settings (TBI-r) only, clustered by subjects. Number of Observations: 399. BC and IC are dummies for the bonus and the incentive contract, respectively. ***, **, and * indicate statistical significance at 0.1%, 1% and 5% level.

Brown and Forsythe variations of Levene's test in the first case the median instead of the mean and in the second case a trimmed mean is used. In all three tests the null hypothesis of equality of variances is rejected with $p < 0.01$. Consequently, the variance in principals' profits is significantly higher for BC than for IC. Hence, risk aversion could be an additional factor affecting contract choice and thus, complementary to profit maximization, explain why IC was chosen more often than BC.

Revealed Effort

Analyzing the results of Table 6 reveals that the size of the fixed wage offer did not affect the decision effort. As neither the coefficient for fixed wage nor one of the interactions of fixed wage with the contract type dummies are significant, this is true for all contract types. This is interesting as many real world contracts still offer fixed wages without any effort dependent compensation part. According to our data, solely raising the fixed wage did not result in more engagement of the agents. However, both variable income parts, i.e. the announced bonus for bonus contracts and the share for incentive contracts show a highly significant positive coefficient. Our interpretation of this result is that if agents were offered a bonus or incentive contract, the size of the announced bonus or share becomes more salient than the fixed wage of such a contract and agents reveal more effort the more they overall profit from such a costly decision. Particularly the latter interpretation is fostered by the positive and highly significant coefficient of the interaction term *IC x Incentive compatible offer*. This means that incentive contracts that offered a share greater or equal to 0.27 (the incentive compatibility threshold), agents were willing to significantly increase their effort by at least 2.6 on average.

Whereas there is no significant difference between the effort levels revealed in TC and BC, the average effort provided in incentive contracts significantly increased by almost eight (columns 1 and 3 of Table 6), which is impressive considering the possible range for e [0,20]. This would mean that agents are willing to provide much more effort if offered an incentive contract, than in a trust contract or bonus contract setting. Focussing on specification 1 and 3 reveals that cheap talk variables, which, from a strict economic point of view, should not have any effect on the revealed effort level, had a significant influence. That is true for both, the *announced bonus* (i.e. a non-binding declaration of a prospective voluntary bonus payment by the principal) and the *demanded effort* (i.e. a non-binding suggestion for the effort level). Whereas compliance to the suggested effort level would have made sense under bonus contracts, in order to persuade the principal to grant a large bonus payment, it does not make sense to send a signal to the principal in an IC setting, as the agent's effort choice finally determines both players' outcome. However, scrutinizing the effect of demanded effort in more detail by adding interaction terms for demanded effort with BC and IC, respectively, changes parts of the results considerably (see columns 2 and 4 of Table 6). The coefficient for demanded effort becomes insignificant, but moreover, none of the interaction terms with demanded effort is significant either, which suits to the above cheap-talk interpretation with respect to demanded effort. Furthermore, controlling for experimental order effects (see **Figure 1)** did not change the regression results significantly. Therefore, the corresponding control dummy was removed from the specifications reported in Table 6.

Bonus payments

The only contract that required an additional decision after the agents' effort levels have been chosen is BC. The principals had to specify which amount [0, 3000] they would like to pay to their agents. A regression analysis that accounted for non-independent observations at subject and session level revealed that the principals indeed took the level of effort into account when choosing the bonus size. Increasing the actual effort level by one unit would ceteris paribus result in a bonus payment raise of almost 36 experimental currency units (ECUs). Furthermore, in specification 1 the coefficient of fixed wage is negative and statistically significant at a 5% level. This result could be interpreted as a trade-off between a high fixed wage and a generous bonus payment such that an increase of the fixed wage by one ECU would reduce the expected bonus payment by 0.3 ECUs. However, the significance vanishes if the non-independence of the data is appropriately taken into account by using a hierarchical linear model (specification 2). The revealed effort level remains as the only significant explanatory variable indicating a 34 ECU bonus increase for one unit of effort increase.

Table 7: Actual Bonus Payments

Dependent Variable:	(1)	(2)
Actual Bonus Payment	Robust Standard Errors	Hierarchical Linear Model grouped by subjects and sessions
Constant	144.561 (158.042)	154.139 (196.942)
Revealed effort	35.951*** (9.954)	34.254** (13.184)
Effort demand exceeded	-187.380 (149.003)	-185.416 (401.613)
Fixed wage	-0.241* (0.117)	-0.220 (0.145)
Announced bonus	-0.002 (0.152)	-0.007 (0.126)
Revealed effort x Effort demand exceeded	12.824 (14.114)	13.531 (24.274)

5. This table reports the coefficients and robust standard errors (in parentheses) of an OLS and a hierarchical linear specification for accepted bonus contracts in repeated contract choice settings (TBI-r) only. Number of Observations: 80. ***, **, and * indicate statistical significance at 0.1%, 1% and 5% level.

5. Concluding Remarks

This article explored how monetary incentives and intrinsic motivation can complement each other in principal-agent settings. In order to consolidate previous findings that contracts with voluntary bonus payments were preferred over enforceable monetary incentives contracts with evidence that positive monetary incentives could achieve efficient yet fair outcomes, we combined Fehr et al.'s (2007) bonus contract with an adapted version of Anderhub et al.'s (2002) revenue sharing contract and added a trust contract as a third option.

In contrast to Fehr et al. (2007) who found that only 10% of the principals chose the enforceable monetary incentivized contract, in our experiment up to 86% of the principals chose such an option. What has been surprising is that in the repeated game (TBI-r) the proportion of incentive contracts was sustained at the same high level as in the one shot game (TBI). One might have expected a higher share of the bonus contract in the repeated game, as reputation effects might have come into play. However, if building trust between principals and agents might have played a role (fixed matching in TBI-r for six rounds), it did not seem to have affected the principals' contract choice behaviour. Furthermore, the incentive contract was the most efficient contract in terms of agents' effort and total revenue and both, in the one shot (TBI) and the repeated (TBI-r) contract choice settings, when the incentive contract was chosen, the distribution of the total surplus was fairer than the profit maximising prediction. The latter result was driven by the principals' choice of providing significantly higher than efficient revenue shares. Thus, share offers were not only incentive compatible, but principals, although being responsive to the monetary incentives, did not only care about efficiency and their personal profit but also showed concerns for fairness. In addition, our experimental results indicate that another explanation of why principals preferred the IC could be attained from the difference in the variance of profits between the IC and the BC. Assuming the principals were risk averse[9], a smaller variance in profits would have implied a higher expected utility and in turn would have favoured the use of IC over the BC.

Overall, this experiment shows that concerns for fairness and reciprocity can go in hand with the use of monetary incentives as long as the correct monetary incentive mechanism is used. Yet, we are still in an early stage of understanding the interaction of monetary incentives and intrinsic motivation and there is still a lot of fascinating and important research to be done. For instance, future research could investigate if the results observed here can be replicated in a multitasking environment or how positively incentivized contracts would be affected by the introduction of risk.

References

Anderhub, V., Gächter, S. & Königstein, M., 2002. "Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment." *Experimental Economics*, 27, pp.25–27.

[9] We did not elicit risk aversion parameters in our experiment.

Bacharach, M., Guerra, G. & Zizzo, D.J., 2007. “The Self-Fulfilling Property of Trust: An Experimental Study.” *Theory and Decision*, 63, pp.349–388.

Benabou, R. & Tirole, J., 2003. “Intrinsic and Extrinsic Motivation.” *Review of Economic Studies*, 70, pp.489–520.

Boly, A., 2010. “On the incentive effects of monitoring: evidence from the lab and the field.” *Experimental Economics*, 14, pp.241–253.

Brown, M.B. & Forsythe, A.B., 1974. “Robust tests for the equality of variances.” *Journal of the American Statistical Association*, pp.364–367.

Dickinson, D. & Villeval, M.-C., 2008. “Does monitoring decrease work effort?” *Games and Economic Behavior*, 63, pp.56–76.

Falk, A., Gächter, S. & Kovács, J., 1999. “Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts.” *Journal of Economic Psychology*, 20, pp.251–284.

Fehr, E. & Gächter, S., 2002. “Do Incentive Contracts Undermine Voluntary Cooperation?” University of Zurich - Working Paper Series (34)

Fehr, E., Kirchsteiger, G. & Riedl, A., 1998. “Gift exchange and reciprocity in competitive experimental markets.” *European Economic Review*, 2921, pp.1–34.

Fehr, E., Klein, A. & Schmidt, K.M., 2007. “Fairness and Contract Design.” *Econometrica*, 75, pp.121–154.

Fehr, E. & Schmidt, K.M., 1999. “A Theory of Fairness, Competition, and Cooperation.” *The Quarterly Journal of Economics*, 114, pp.817–868.

Fehr, E. & Schmidt, K.M., 2007. “Adding a Stick to the Carrot? The Interaction of Bonuses and Fines.” *The American Economic Review*, 97, pp.177–181.

Fehr, E. & Schmidt, K.M., 2004. “Fairness and Incentives in a Multi-Task Principal-Agent Model.” , University of Munich, Munich Discussion Paper No. 2004-8.

Gneezy, U. & Rustichini, A., 2000. “Pay Enough or Don’t Pay at All.” *The Quarterly Journal of Economics*, 115, pp.791–810.

Güth, W., Schmittberger, R. & Schwarze, B., 1982. “An experimental analysis of ultimatum bargaining.” *Journal of Economic Behavior and Organization*, 3, pp.367–388.

James, H.S., 2005. “Why did you do that? An economic examination of the effect of extrinsic compensation on intrinsic motivation and performance.” *Journal of Economic Psychology*, 26(4), pp.549–566.

Lazear, E.P., 1986. “Incentive Contracts.” National Bureau of Economic Research - Working Paper Series (1917).

Levene, H., 1960. “Robust Tests for Equality of Variances.” In I. Olkin et al., eds. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling.* Stanford, California: Stanford University Press, pp. 278–291.

Murdock, K., 2002. “Intrinsic Motivation and Optimal Incentive Contracts.” RAND *Journal of Economics*, 33, pp.650–671.

Prendergast, C., 2008. “Intrinsic Motivation and Incentives.” *American Economic Review*, 98, pp.201–205.

Appendix A

Parameters of the Experimental Implementation

Three different contracts types are used in this experiment, the trust contract (TC), the bonus contract (BC) and the incentive contract (IC).

The agent’s profit in the case of an *incentive contract* is defined as:

$$\mathbf{P_{IC}^{A} = F + S * R(e)} \quad (1)$$

The agent’s profit in the case of a *trust contract* is defined as:

$$\mathbf{P_{TC}^{A} = F} \quad (2)$$

The agent’s profit in the case of a *bonus contract* is defined as:

$$\mathrm{P_{TC}^{A} = F + B} \quad (3)$$

The *Total Revenue* is given by:

$$R(e) = 150 * e \tag{4}$$

The cost of effort is a strictly increasing and convex function in effort:

$$C(e) = e + e^2 \tag{5}$$

With:

$P^{A}_{TC,BC,IC}$ ... Agent’s profit in the case of TC, BC and IC, respectively

F ... Unconditional fixed wage

$F \in \{0,1,\dots,2999,3000\}$

$R(e)$ ... Revenue

S ... Relative share of the Revenue that is transferred to the agent in the incentive contract

$S \in \{0.00, 0.01,\dots,0.99,1.00\}$

B ... Optional bonus paid to the agent in a bonus contract

$B \in \{0,1,\dots,2999,3000\}$

e ... Effort level revealed by the agent

$e \in \{0,1,\dots,19,20\}$

The game theoretic solution

Given the above parameters the *participation constraint*, i.e. the constraint that has to be met in order to make any contract offer monetarily beneficial is:

$$T(e) - C(e) \geq 0 \tag{6}$$

or stated differently

$$T(e) \geq C(e)$$

Where, $T(e)$ is the *transfer* the principal needs to provide to the agent as compensation for exerting effort. The nature of that transfer depends on the type of contract that will be chosen from the principal. Thus agents should only accept a contract if (6) is met.

The Principal's profit is defined as

$$P^P = R(e) - T(e) \qquad (7)$$

Where $R(e)$ is the total revenue generated and $T(e)$ is the transfer to the agent.

Thus the principal wants to:

$$\max\,(R(e) - T(e))$$

Given that the agent would accept any contract that satisfies (6) the minimum amount that has to be transferred to the agent has to be equal to $C(e)$. Thus if

$$T(e) = C(e)$$

The principal's maximization problem becomes

$$\max\,(R(e) - C(e)) \qquad (8)$$

Inserting the actual parameters used in the experiment results in

$$\max\,(150e - e - e^2)$$

Maximizing by e results in

$$e = 74.5$$

Thus the optimal effort level would be 74.5. As the experimental parameters only allow

$e \in \{0,1,\ldots,19,20\}$, the maximisation problem in (7) has a *corner solution*[10] of $e^* = 20$.

Having identified the participation constraint and the profit maximising effort level, the following step is to show why, given the assumption that both the principal and the agent are rational and narrowly self-interested, the only contract that can satisfy the *incentive compatibility constraint* is the IC.

Any contract is deemed to only be *incentive compatible* if:

[10] The decision the maximisation problem to have a corner solution has been made deliberately under the suspicion that will be easier for subjects in the role of principals to identify e* if that is a corner than an interior point. In other words, the choice for a corner solution was made to reduce complexity to an already highly complex design from the perspective of the principal. This choice though bears the cost that will be harder to test if the principals had correctly identified that e* is the optimum effort level or if they chose it ad hoc simply following a rule of thumb such as the more the better. Nevertheless, the use of corner solutions has been a standard approach to experiments which investigated contract design and social preferences.

$$\forall \mathrm{e}: \mathrm{T}(\mathrm{e}^*) - \mathrm{C}(\mathrm{e}^*) \geq \mathrm{T}(\mathrm{e}) - \mathrm{C}(\mathrm{e}) \tag{9}$$

Inequality (9) implies that the agent's profit from exerting effort level e^* (which is maximizing the principal's profit) should be greater or equal to the profit that results from exerting all possible effort levels e.

The following three sections examine incentive compatibility for the incentive, trust and bonus contracts respectively, by substituting $T(e)$ by the specific transfer definitions of each of the three contracts.

<u>The incentive contract:</u>

Replacing $T(e)$ with the incentive contract specific transfer of $T_{IC} = F + S * 150e$ results in:

$$\forall \mathrm{e}: \mathrm{F} + \mathrm{S} * 150\mathrm{e}^* - \mathrm{e}^* - \mathrm{e}^{*2} \geq \mathrm{F} + \mathrm{S} * 150\mathrm{e} - \mathrm{e} - \mathrm{e}^2 \tag{10}$$

Given that the agent would exert an effort greater than zero if $P(e^*) \geq P(e)$ is satisfied, the agent would, as a worst case accept, $P(e^*) = P(e)$. Consequently, in order to calculate the minimum share of the total revenue that has to be provided to the agent in order to make the incentive contract incentive compatible, the profit maximization problem for the agent could be written as

$$\mathrm{P}_{\mathrm{IC}}^{\mathrm{A}} = \mathrm{F} + \mathrm{S} * 150\mathrm{e}^* - \mathrm{e}^* - \mathrm{e}^{*2} \tag{11}$$

Maximizing (11) with respect to e^* leads to

$\mathrm{S} * 150\mathrm{e}^* - 1 - 2\mathrm{e}^* = 0$ and

$$\mathrm{S} = \frac{2\mathrm{e}^* + 1}{150}$$

Inserting the above calculated effort level $e^* = 20$ and solving for S, finally provides the minimum share S. Thus,

$$\mathrm{S} = 0.27\dot{3} \tag{12}$$

Thus, the incentive contract is incentive compatible for any value of S $\geq 0.27\dot{3}$.

With S=S* the consequent profits for the principal and the agent are respectively: $P^P = 5{,}220$ECU[11] and $P^A = 3{,}360$ECU.

The trust contract:

Replacing $T(e)$ with the trust contract specific transfer of $T_{TC} = F$ results in the incentive compatibility constraint for all trust contracts:

$$\forall e\colon \mathrm{F} - \mathrm{C}(\mathrm{e}^*) \geq \mathrm{F} - \mathrm{C}(\mathrm{e}) \tag{13}$$

Which can be restated as:

$$\mathrm{C}(\mathrm{e}^*) \leq \mathrm{C}(\mathrm{e}) \tag{14}$$

Because of (5) the only value of e^* that satisfies equation (14) is $e^* = 0.$ Therefore, there exists no feasible incentive compatible trust contract for $e^* > 0$.

The bonus contract:

Replacing $T(e)$ with the bonus contract specific transfer of $T_{TC} = F + B$ results in the incentive compatibility constraint for all bonus contracts:

$$\forall e\colon \mathrm{F} + \mathrm{B} - \mathrm{C}(\mathrm{e}^*) \geq \mathrm{F} + \mathrm{B} - \mathrm{C}(\mathrm{e}) \tag{15}$$

Rewriting leads to:

$$\mathrm{C}(\mathrm{e}^*) \leq \mathrm{C}(\mathrm{e}) \tag{16}$$

This is identical to the result obtained for the trust contract. Therefore, it has been shown that economic theory predicts that under the assumption of selfish rational profit maximizing individuals no agreement can be reached between a principal and an agent in neither the trust nor the bonus contracts. From the results obtained above, it is clear that the only contract that can satisfy both the incentive compatibility and the participation constraints is the incentive contract IC. Consequently, the game theoretic solution that is expected in the TBI and TBI-r game(s) is that IC should dominate both BC and TC.

[11] ECU stands for Experimental Currency Units

Appendix B - Instructions

General Instructions

Welcome to our experiment! Please read the following instructions carefully. Reading these instructions carefully could earn you a significant amount of money. If you face any difficulties understanding any part of the instructions please raise your hand and an experimenter will come to assist you. All the money that you will earn during this experiment will be paid to you in cash at the end of this experiment.

No talking is allowed through the experiment. Please switch off your mobile phones.

Experiment Overview

Each participant is assigned randomly the role of either the employer or the employee. There is a note on your desk clarifying your role. Communication between the two will be via the computer**.** The experiment is **anonymous**; this means that you will not know with of the other participants you are interacting. Interaction will be through contracts. A contract is an offer by the employer to the employee for offering a value of effort. The details are discussed below.

The experiment consists of **3 practice stages**, and **5 real stages**. In the 3 practice stages every employer is matched with the **same** employee. In the real stages, the employer will be matched with a **different** employee in every stage who will also be different from the one he/she encountered in the practice stages. The practice stages are to help you familiarise with the procedure of the experiment and your choices will not affect your earnings. The following five ‘real’ stages form the main body of the experiment and your choices will affect your final earnings. The 5 real stages consist in total of 14 rounds. At the end of the experiment the earnings you made from one of these rounds are randomly chosen by the computer and are added to your show up fee.

For attending this experiment you will be given a show up fee of £3. In the experiment you will be using an experimental currency called ECU. In the end of the experiment the ECU you have earned during the experiment will be exchanged at the exchange rate of: **250ECU = £1.**

For example, 500ECU=£2, 1000ECU=£4, 25ECU= £0.10, 3000= £12.

At the start of each stage a new set of instructions is given to you which, will explain the process of the stages that is starting and accompany the instructions for the following stages.

Stage 1: Contract 1 (practice)

In this round the employer has to decide the size of a **fixed wage** that he/she wants to pay the employee, and set a **suggested effort** level. The fixed wage can range between 0 and 3000 and the suggested effort from 0 to 20. Both the fixed wage and suggested effort are received by the employee **before** he/she decides an effort level.

The employee has to choose an effort level which, **for every unit of effort the employee spends, you earn 150ECU**; we call this *total revenue*. The **total revenue=150 x effort** (see Table 8 below).

At the start of every round both employer and employee are given a capital of 3000ECU this money is for you to use within the experiment and are added to your earnings for the round.

<u>There are three key elements you need to note:</u>

Firstly, for every unit of effort the employee spends, it has a subsequent ECU cost to him. The exact cost of ECU for every unit of effort along with other important information is shown in Table 8 which is handed in a separate sheet.

Secondly, the suggested effort of the employer is only a suggestion. The employee is not bound to that suggestion but he/she is free to choose any effort level within the given range of 0 to 20.

Thirdly, the fixed wage is paid upfront (i.e. before the employee decides an effort level).

How earnings are calculated

For the employer his/her earnings are the capital plus the total revenue generated by the employee's effort minus the fixed wage he/she paid. In other words:

Employer's Profit= Employer capital + Total revenue – fixed wage

In the case of the employee, his/her profits are his/her capital plus the fixed wage minus the cost of effort. In other words:

Employee's Profit= Employee capital + fixed wage – cost of effort

The process of the stage is the following:

0. Before the stage starts, there are four multiple choice quizzes to check that you understood what your earnings will be according to your choices.

1. The employer chooses the fixed wage and suggest an effort level to the employee.

2. Afterwards, the employee has been informed of the offered contract, he/she has to decide either to accept or reject the contract.

3. If the employee rejects the contract the stage finishes. If he accepts the contract, he receives the offered fixed wage and decides an effort level.

4. Once the employee has decided an effort level, the computer calculates and informs both participants of their profits.

Some Examples

Example 1: Assume the employer decides to offer a fixed wage of 500ECU, sets suggested effort to 20 and the employee decides to accept the offer and offer an effort level of 20. What would the profits of the employer and employee be?

Answer: By looking on Table 8 we can see that the total revenue for 20 units of effort is 3000 ECU. So the profits for the employer are 3000ECU (the total revenue) plus the employer capital of 3000ECU minus 500ECU (the fixed wage), therefore 5500 ECU. For the employee the profits are his/her capital of 450ECU plus 500ECU (the fixed wage) minus the cost for his effort which is 420ECU (see Table 8), therefore 3530 ECU.

Example 2: Assume like before that the employer offers a fixed wage of 500ECU and sets a suggested effort of 20 and the employee decides to accept the offer and offer an effort level of 0. What would the profits of the employer and employee be?

Answer: In this case the total revenue is 0ECU. The employer receives only his capital of 3000 which from 500ECU are subtracted (the fixed wage he/she paid) hence he/she earns 2500 ECU. The employee earns 500ECU (the fixed wage) plus his/her capital of 3000ECU therefore he/she earns 3500 ECU.

Stage 2: Contract Type 2 (practice)

Round 2 is identical to round 1 with the only exception that now the employer can also announce a **bonus** to the employee. When the employer offers the contract, except of the fixed wage, he/she can also announce a bonus. However, the bonus announcement is non-binding. That is, after the earnings for both of you are realised, the employer is free to decide if he/she wants to pay a bonus or not and if so of what size.

Summing up, the employer has to pay a **fixed wage upfront**, announce a non-binding **bonus** and suggest **an effort level**. After the employee decides an effort level, the employer has to decide the size of the bonus he/she wants to pay. Both fixed wage and bonus can range from 0ECU to 3000ECU but also the sum of the two (fixed wage and bonus) cannot exceed 3000ECU.

The process of the stage is the following:

0. Before the stage starts, there are four multiple choice quizzes to check that you understood what your earnings will be according to your choices.

1. The employer chooses the size of the **fixed wage**, the size of the **announced bonus** and **suggests an effort level** to the employee.

2. After being informed of the offered contract the employee has to decide either to accept or reject the contract.

3. If the employee rejects the contract the stage finishes. If he/she accepts the contract, receives the offered fixed wage and decides an effort level.

4. After the employee had decided an effort level, the computer calculates and informs both employer and employee their profits. At this point the employer will be asked if he/she wants to pay a **bonus** and if so, of what size. Depending on the employer’s choice the computer recalculates and informs both of you for your final profits for this stage.

How earnings are calculated

For the employer, his/her earnings are the capital plus the total revenue generated by the employee’s effort minus the fixed wage and minus any bonus he/she paid. In other words:

Employer’s Profit = Employer capital + total revenue – fixed wage – bonus

In the case of the employee, his/her earnings are the employee capital plus the fixed wage plus any bonus minus the cost of effort. In other words:

Employee’s Profit= Employee capital + fixed wage + bonus – cost of effort

Some Examples

Example 1: Assume the employer decides to offer a fixed wage of 500 ECU, announces a bonus of 500ECU and sets suggested effort to 20. The employee decides to accept the offer and offer an effort level of 20. Then the employer gets informed about the total revenue and decides to pay a bonus of 400 ECU. What would the profits of the employer and employee be?

Answer: By looking at Table 8 we can see that the total revenue for 20 units of effort is 3000 ECU. So the profits for the employer are his/her capital of 3000 plus 3000ECU (the total revenue) minus 500ECU (the fixed wage), minus the bonus of

400ECU, therefore 2900 ECU. For the employee the profits are his/her capital of 3000ECU plus 500ECU (the fixed wage) plus the bonus of 400ECU minus the cost for his effort which is 420ECU (see Table 8), therefore 3480 ECU.

Example 2: Assume the employer decides to offer a fixed wage of 700 ECU, announce a bonus of 500ECU and sets suggested effort to 20. He observes a total revenue of 1500 ECU. i) What was the effort level that the agent chose? ii)If the employer decides to pay a bonus of 0, what would the profits of the employer and employee be?

Answer: i) The employer by looking on Table 8 can see that a total revenue of 1500 ECU corresponds to an effort level of 10. ii) For a total revenue of 1500 ECU, the employer earns his/her capital of 3000 ECU plus 1500 (the total revenue) minus the fixed wage of 700 hence his/her profits are 3800 ECU. The employee earns his her capital of 3000 ECU plus 700 ECU (the fixed wage) minus the cost of effort for 10 units of effort which is 110 ECU. Thus, the employee earns 3590 ECU.

Stage 3: Contract Type 3 (practice)

In this stage the employer instead of a bonus he/she can offer a **share of the total revenue** to the employee. This offer is binding. That is, that as long as the employer has offered a share of the total revenue to the employee he/she cannot change the offer.

For example, a value of 0.09, 0.54 or 0.92 will correspond to 9%, 54% or 92% of the total revenue being given to the employee.

Like before you can also offer a fixed wage, between 0 and 3000, and again you have to suggest an effort level.

The process of the stage is the following:

0. Before the stage starts, there are four multiple choice quizzes to check that you understood what your earnings will be according to your choices.

1. The employer chooses the size of the fixed wage, the size of the share of total revenue he/she wants to offer, and suggests an effort level to the employee.

2. After being informed of the offered contract, the employee decides either to accept or reject the contract.

3. If the employee rejects the contract the stage finishes and you move to the next stage. If he/she accepts the contract he/she receives the offered fixed wage and decides an effort level.

4. After the employee had decided an effort level, the computer calculates the total revenue, allocates it between the employer and the employee according to the size of the share that each of them holds, and informs both about their final profits.

How earnings are calculated

For the employer, his/her profits are the employer capital, the total revenue generated by the employee's effort minus the fixed wage, minus the share of the total revenue he/she offered to the employee. In other words:

Employer's Profit= Employer capital + total revenue – fixed wage – share * total revenue

In the case of the employee, his/her profits are the employee capital, plus the fixed wage plus the share on the total revenue that has been offered to him/her, minus the cost of effort. In other words:

Employee's Profit= Employee capital + fixed wage + share * total revenue – cost of effort

Some Examples

Example 1: Assume the employer decides to offer a fixed wage of 200ECU, offer a share of 0.2, and set suggested effort to 15. The employee decides to accept the offer and offer an effort level of 20. What would the profits of the employer and employee be?

Answer: By looking on Table 8 we can see that the total revenue for 20 units of effort is 3000 ECU. So the profits for the employer are 3000ECU (the total revenue) minus 100 ECU (the fixed wage), minus the share (0.2 x 3000 =600), therefore 2300 ECU plus the employer capital of 3000 ECU hence 5300 ECU. For the employee the profits are the employee capital of 3000 ECU, plus 100 ECU (the fixed wage) plus the share of 600 ECU minus the cost for his effort which is 420 ECU (see Table 8), therefore, 3280 ECU.

Example 2: Assume the employer decides to offer a fixed wage of 0ECU, offer a share of 0.6, and set suggested effort to 20. The employee decides to accept the offer and offer an effort level of 18. What would the profits of the employer and employee be?

Answer: By looking on Table 8 we can see that the total revenue for 18 units of effort is 2700 ECU. So the profits for the employer are 2700 ECU (the total revenue), minus the share (0.6 x 2700 =1620) plus his capital of 3000 ECU, therefore 4080 ECU (2700-1620=1080 +3000). For the employee the profits are the share of 1620ECU minus the cost for his effort which is 342 ECU (see Table 8) plus his/her capital of 450, hence, 4278 ECU.

Note: to make your calculations easier recall that a percentage of say 2%, 20%, 100%, its equal to 0.02, 0.2 and 1 respectively.

Stage 4: Contract Type 1

From now on your choices affect your earnings. You should keep in mind the clock on the top right side of the screen and comply with the time constraints

This stage is the same as stage 1 but this time your choices affect your earnings. For how earnings are calculated or for the procedures of the stage you should recall on the instruction sheet that was given to you at the start of stage 1.

Reminder

Type 1: Fixed Wage

Stage 5: Contract Type 2

This stage is the same as stage 2 but this time your choices affect your earnings. For how earnings are calculated or for the procedures of the stage you should recall on the instruction sheet that was given to you at the start of stage 2.

Reminder

Type 2: Fixed Wage + Bonus

Stage 6: Contract Type 3

This stage is the same as stage 3 but this time your choices affect your earnings. For how earnings are calculated or for the procedures of the stage you should recall on the instruction sheet that was given to you at the start of stage 3.

Reminder

Type 3: Fixed Wage + Share

Stage 7: Choice among the 3 Contracts

In this stage the employer is given the option to choose between the three possible contracts that you experienced before. Therefore, he/she firstly has to choose which of the three contracts he/she want to use and the rest of the stage follows exactly as in the corresponding stage you participated earlier.

Reminder

Type 1: Fixed Wage

Type 2: Fixed Wage + Bonus

Type 3: Fixed wage + Share

The process of the stage is the following:

1. The employer chooses one of the three contracts.
2. The remaining procedure is identical to the corresponding contract you practiced with before.

For any queries on how earnings are calculated see the instructions that were provided to you.

Stage 8: Choice between the 3 Contracts - repeated interaction

This stage is identical to stage 4 with only difference that is consisted of 6 rounds in which you are paired with the same participant. In each round the employer has to choose one of the three contracts and according to his/her choice the stage continues.

Note: At the start of every round both the employer's and employee's capitals are refreshed. In addition, if a contract is rejected the stage is not finished but you move to the next round of the stage.

Reminder

Type 1: Fixed Wage

Type 2: Fixed Wage + Bonus

Type 3: Fixed wage + Share

Effort Level	Cost of Effort	Total Revenue
0	0	0
1	2	150
2	6	300
3	12	450
4	20	600
5	30	750
6	42	900
7	56	1050
8	72	1200
9	90	1350
10	110	1500
11	132	1650
12	156	1800
13	182	1950
14	210	2100
15	240	2250
16	272	2400
17	306	2550
18	342	2700
19	380	2850
20	420	3000

Table 8: Effort levels, Cost of Effort, and Total Revenue

Employer Capital: 3000 ECU

Employee Capital: 3000 ECU

Chapter 2: An experimental investigation of the tenuous trade-off between risk and incentives.

1. Introduction

Most of the work in principal-agent theory has focused on how to design contracts which incentivise agents to provide the optimal, for the principal, effort level. Central to this work is the Incentive Intensity Principle (Holmstrom and Milgrom, 1987). The incentive intensity principle describes the optimal incentive intensity of the contract the principal offers to an agent when effort is not directly observable and contractible but instead the observed output is subject to a stochastic random factor. Under the assumption that the principal is risk neutral and the agent is risk averse, a key result of the model is that there is an inverse relationship between incentive intensity and the variance of the stochastic random factor which affects the final output. However, the evidence on the relationship between incentive intensity and risk are mixed (Prendergast, 1999). That led some scholars to describe this relationship as 'tenuous' and to attempt to provide alternative justifications of why the expected relationship is not observed in the studies (i.e. Prendergast, 2002). In this paper we present an experiment which tests the relationship between incentive intensity and risk.

Conducting an experiment to test this relationship is important because there is no previous research which tested this relationship in a lab. Testing this relationship in a lab has two significant advantages: Firstly, in the lab there is enhanced control which allows obtaining more and better information on of the parameters of the model, and secondly, allows ruling out alternative explanations (i.e. other parameters) of why the relationship observed is weak. This way this experiment will not only provide an extra piece of evidence on the relationship between risk and incentive intensity experimentally, which has not been done before, but also indirectly test whether these alternative explanations may be the reason why the observed relationship is tenuous.

The remainder of this chapter is structured in five sections: Section 2 provides a literature review; Section 3 presents the theoretical framework and baseline experimental predictions; Section 4 explains the experimental design. Section 5 presents the results, Section 6 discuss the results of the experiment and section 7 concludes. The experimental instructions as well as proofs for the theoretical prediction are provided in the appendix.

2. Literature Review

Central in principal-agent theory is the provision of incentives using pay for performance schemes. The key intuition is that pay for performance contracts (such as piece rates) are useful when there is little or no noise in between output and effort. If there is a lot of noise and effort cannot be distinguished by output two problems arise. Firstly, the piece rate is no longer a good indicator of effort. Secondly, if the agent is risk averse the principal would have to provide insurance through higher wages in order to ensure the agent's participation to the contract. The Incentive Intensity Principle (henceforth, IIP) developed by Holmstrom and Milgrom (1987) examines this problem. According to the IIP the optimal intensity of incentives depends on four factors: i) the incremental profit generated by each extra unit of effort, ii) the risk tolerance of the agent, iii) the level of risk in the environment, and iv) the responsiveness of effort to incentives. One of the key results of the IIP is that risk and incentive intensity have a negative relationship.

A recent development in the theoretical literature on incentive contracts has been provided by Englmaier and Wambach (2010) expanded the moral hazard model to incorporate the inequity aversion model of (Fehr and Schmidt, 1999). With the introduction of inequity aversion, when there is risk associated with output and the agents are risk averse the slope of the optimal scheme is below 1/2 in the First Best and tends towards the 1/2 the more inequity averse the agent is. However, in the Second Best, when the principal needs to incentivise the agent there are no clear cut predictions. This is because there are three forces in action: Inequity aversion which pushes towards an equal split; risk aversion which pushes towards a slope of 0 (i.e. flat wages); and the need to provide incentives to the agent where the slope is maximal at 1. As a result, deriving predictions for incentive intensity under inequity aversion is a dubious task as one needs measures for both risk aversion, inequity aversion while also taking into account incentive compatibility.

A relevant literature for this chapter would also be the research on risk elicitation procedures. Harrison and Rutström (2008) provide an extensive and recent review on the most commonly employed measures of risk aversion. There are five general types of measures of risk aversion which have been employed by economists. These are: the multiple price list (MPL), in which the subjects are given an ordered array of binary lottery choices to make all at once; the random lottery pairs (RLP), in which the subjects choose one of the lotteries in each pair and face multiple pairs in sequence; the ordered lottery selection (OLS), in which the subjects select one lottery from an ordered set; the Becker-DeGroot-Marschak auction (BDM), in which the subjects are asked to choose a minimum certainty equivalent for a lottery that they have been endowed with; and the trade-off (TO) design which is a hybrid of the others (Harrison and Rutström, 2008). The most notable application of the MPL is the Holt and Laury (2002) questionnaire, the main weaknesses of the MPLare (a) that provides an interval estimate rather than a specific point estimate and (b) it can violate monotonicity, however its simplicity and relatively transparent procedure have made it very popular among experimental economists (Harrison and Rutström, 2008). The RLP design (e.g. Hey and Orme, 1994) is easy for subjects to understand, however it is not possible to directly infer risk attitudes and requires some form of estimation. The OLS's (e.g. Eckel and Grossman, 2002, 2005) main disadvantage is that the use of a certain amount as a safe option may be perceived as a reference point to identify gains and losses. Lastly the BDM is very complex for subjects to understand and the TO is not incentive compatible (Harrison and Rutström, 2008). Given the above we had concluded for the purposes of this experiment to measure risk aversion using the Holt and Laury (2002) risk elicitation questionnaire while restricting for monotonicity.

The empirical evidence on the trade-off between risk and incentives are mixed (Prendergast, 1999). Prendergast (2002) provides a summary of the empirical findings of 28 papers. Out of the 28 papers only 4 found a negative relationship (i.e. Brown, 1990; Garen, 1994; Bushman *et al.*, 1996) as theory predicts, while 13 found no statistically significant relationship (i.e. Lambert and Larker, 1987, Aggerwal and Samwick, 1998) and 11 found a positive relationship (i.e. Norton, 1988; Lafontaine, 1992; Core and Guay, 1999). However, there have been no experimental investigations of the trade-off between risk and incentives.

Anderhub et al. (2002) conducted a simple principal-agent experiment in which the employer had to determine what share of the gross revenue wanted to offer to the agent. In their experiment there was no risk associated with output. The principals designed contracts which were in most of the cases incentive compatible, aimed at efficiency, and satisfied the participation constraint, and additionally were in general more generous than what standard economic theory predicts (Anderhub *et al.*, 2002). On the other hand, the agents who received generous contracts responded by supplying effort levels above their best reply levels compared to agents that received 'unfair' contracts. Furthermore, offers that were deemed 'unfair' by agents had been rejected (Anderhub *et al.*, 2002).

According to the linear agency model the optimal effort choice of the agent depends on the marginal cost of effort and is unrelated to the noise in the performance measure. Sloof and van Praag (2008) run an experimental study to test this prediction and compare with expectancy theory, a theory developed by psychologists which predicts a negative relationship between effort and noise in the performance measure. In contrast to this study which focuses on the optimal choice of incentive intensity *(β),* Sloof and van Praag (2008) are focused on the optimal effort choice. For this reason and to reduce complexity they have opted out for the role of the principal to be determined exogenously. In particular the subjects were given the role of sales representatives who had to allocate effort between two different tasks /regions in each of 30 rounds. The earnings of the subjects were based on a performance pay measure which was split in three parts: A fixed wage, a share of 50% of the overall sales for region A and 50% of the overall sales in region B. Overall sales then depended on the effort level of the participants and the noise terms for each region which would differ every five rounds. Their findings are in line with the linear agency model as the results suggest that effort levels were invariant to the distributions of the noise terms.

3. Principal-Agent Problem

In this section we describe the game theoretic solutions firstly for the case where there is no stochastic variance on output, therefore effort is observable but not contractible, and secondly when output is the sum of effort and a stochastic random factor, therefore the principal cannot distinguish which part of the output is due to the agent's effort and which part is due to the stochastic random factor.

When there is no risk, the revenue of production depends on the agent's effort level e such that$R(e) = 50 * e$. For providing effort, the agent bears a cost of $C(e) = e^2$ with $e \in \{5,6,\ldots,10\}$. The principal jointly decides on a fixed wage $F \in \{50, 51,\ldots,200\}$ and a piece rate $\beta \in \{5, 6,\ldots,40\}$ that specifies how much Experimental Currency Units the agent will receive for each unit of effort. The agent's payoff is therefore given by $\mathrm{F} + \beta * e - C(e)$. Conversely, the principal's payoff is $\mathrm{R(e)} - \mathrm{F} - \beta * e$. Assuming that both principals and agents are profit maximisers, the game theoretic solution predicts the principal to offer a piece rate of 20 and a fixed wage of 50 (due to the parameterization of the experiment) and the agent to exert an effort level of 10[12]. The consequent profits (after taking into account their endowments) for the principal $P^P = 450$ and $P^A = 350$ for the agent.

When there is risk in the environment, the piece rate in output is subject to stochastic variance (*V*) which is assumed to be normally distributed with a mean of zero. Assuming the principal is risk neutral and the agent risk averse the principal faces a trade-off between incentivising the agent and providing him with insurance for the variance in payoffs that is created due to the stochastic random factor. Holmstrom and Milgrom (1987) have shown that the optimal incentive intensity (i.e. piece rate) is given by:

$$\beta^* = \frac{R'(e)}{1 + rVC''(e)}$$

[12] For all derivations see appendix B.

Where r is the coefficient of absolute risk aversion of the agent. Given that V in our experiment was set at 2.5 and the rest of the factors remained unchanged the optimal β^* is given by:

$$\beta^* = \frac{50}{1+5r}$$

We elicited the coefficient of absolute risk aversion for all subjects by using the Holt and Laury (2002) questionnaire. In order to generate a benchmark for our analysis we calculated an average $\bar{r}$ from all subjects in our experiment.[13] The average coefficient of absolute risk aversion from all subjects who participated in our experiment was $\bar{r}$ = 0.549. After inserting $\bar{r}$ in equation 1 it yields the optimal β^*= 13.35. Given the optimal β^* the optimal effort level for the employee is 7. The consequent expected profits (after taking into account their endowments) for the principal $P^P = 406$ and $P^A = 294$ for the agent.

4. Experimental Design

The experiment was conducted at the University of East Anglia with 360 participants. The participants were mostly students with a variety of different backgrounds. The experiment was computerised in Z-Tree (Fischbacher, 2007). The experiment employed a fictional currency, called ECU, which was converted to pounds at the end of the experiment at the rate of £0.02 per ECU. Each session lasted approximately 80 minutes and the subjects earned on average £9.60 (approximately 15.12 US dollars), including a show-up fee of £2.00. Each session consisted of 10 rounds where the first three rounds where for practice (i.e. these rounds did not affect the subjects' earnings).

At the end of the 10 rounds the subjects had to complete the Holt and Laury (2002) risk elicitation questionnaire. In addition, after the completion of the Holt and Laury (2002) risk elicitation questionnaire the subjects had to complete, two

[13] Note that the model assumes that the principal is aware of the exact value of the coefficient of absolute risk aversion for each agent.

non incentivised psychology questionnaires which measure risk taking and risk perception developed by Blais and Weber (2006). In the end of the experiment one of the seven rounds was chosen randomly and was paid to subjects in cash along with any additional earnings from the Holt and Layry (2002) task. Earnings were paid privately and anonymously. A random matching procedure was implemented at the start of each round to control for reputation effects. Subjects were not allowed to participate in more than one session. A positive frame of employer/ employee was adopted, instead of Type A and Type B, as context can be useful to enhance understanding (see Cooper and Kagel 2003 and 2009). In addition, both the employer /employee frame (eg. Fehr, et al. 1998) and the buyer/ seller frame (eg. Fehr and Gachter 2002, and Fehr et al. 2007), have been previously used in the context of the gift exchange with no qualitative differences between the two frames.

The experiment consisted of three treatments: No Risk (NR), Risk (R), and Both in Risk (BR). We ran 30 sessions in total with 12 subjects in each session. The subjects were split evenly as employers or employees and maintained the same role throughout the experiment.

The instructions were common for both employers and employees. Before the start of each session and after the subjects had read the instructions they had to answer control questions to ensure that all subjects have understood the instructions. If a subject provided the wrong answer in any question a detailed explanation was provided in his or her computer screen.

All three treatments were identical in every aspect apart from how in each round risk affected the profit functions of the employers and employees. In the NR treatment, the employer had to offer an employment contract to the employee requesting him or her to exert a level of effort. In the employment contract, the employer had to specify the size of the fixed wage, piece rate and suggest an effort level. The fixed wage could range between 50 and 200, the piece rate between 5 and 40, and the suggested effort level between 5 and 10.[14] Afterwards the employee

[14] In order to ensure that no subject made any losses due to the variance of the random factor, we had decided that there would be a minimum effort level and consequently a minimum wage. To avoid creating any potential cues regarding which payment mechanism the employer should use. We split the minimum wage equally between the fixed wage and piece rate.

had to decide whether to accept or reject the contract offer. If the contract was rejected the round finished and the subjects would earn only their endowments which were 200 ECU. If the employee accepted the contract then he or she had to decide an effort level between 5 and 10. Effort was costly and the cost of effort was given by the function *C(e) =*e^2, where e stands for effort. The total revenue for the employer was given by TR*=50e*.[15] After the subject decided an effort level, profits were given by the following functions:

$$\pi_{Employer} = Endowment + 50e - [Fixed\ wage + (Piece\ Rate) * e]$$

$$\pi_{Employee} = Endowment + Fixed\ wage + (Piece\ Rate) * e - C(e)$$

In the R treatment the profit function of the employee was altered to incorporate the risk associated with the incentive measure (the piece rate). According to the theory the random factor is assumed to generate noise in the performance measure not allowing the principal to directly observe the effort choice of the agent.[16] Given that the principal is assumed to be risk neutral and the random factor is assumed to be normally distributed with a mean of zero, he or she is expected not to be affected by the associated risk in the total revenue nor on its impact on the piece rate. The main concern for the principal is that the random factor dilutes the incentives generated by the piece rate for the risk averse agent. According to the model the principal is assumed to be risk neutral as he or she is able to diversify the associated risk. However, given the principal is assigned only one employee this assumption would not be justifiable. In order to overcome this problem and ensure the principal can act as risk neutral the risk component was removed from his or her profit function. As a result the experimenter acts as an insurer for the principal allowing him to act as if he or she was risk neutral. This allowed us to rule out any effects from the principal being risk averse which would

[15] In other words the employers would receive 50 ECU for each unit of effort that was provided by the employee.

[16] The random factor was presented to the subjects in the form of a table in which each possible value that x could take was assigned a respective probability. The table can be found in the attached copy of instructions in Appendix A.

deviate from a key assumption of the model. Therefore, the profit function of the employer was held unchanged, whereas, the piece rate that was paid to the agent was formulated by the sum of the effort and the random factor (x).

$$\pi_{Employer} = Endowment + 50e - [Fixed\ wage + (Piece\ Rate) * e]$$

$$\pi_{Employee} = Endowment + Fixed\ wage + (Piece\ Rate) * (e + x) - C(e)$$

The BR treatment was run as a control treatment in order to double check whether imposing risk neutrality to the principal does have an effect in his or her behaviour. The profit function of the agent remained as in the R treatment. Whereas the principal in this treatment was also subject to risk.

$$\pi_{Employer} = Endowment + 50(e + x) - [Fixed\ wage + (Piece\ Rate) * (e + x)]$$

$$\pi_{Employee} = Endowment + Fixed\ wage + (Piece\ Rate) * (e + x) - C(e)$$

As we have shown in the previous section that the model suggests that the optimal piece rate β will be 20 in the NR treatment and 13.35 in the R treatment. This is because the employer partially ensures the employee.

Hypothesis 1: The piece rate will be larger in the R treatment than in the NR treatment.

Hypothesis 2: The piece rate offered by the principals in the NR treatment will be on average equal or larger to 20 ECU per unit of effort.

Hypothesis 3: The piece rate offered by the principals in the R treatment will be on average equal or larger to 13.35 ECU per unit of effort.

The agent's effort level is expected to depend only on the piece rate (and not on the fixed wage or the noise in the performance measure) offered by the principal as it formulates the incentive constraint of the agent.

Hypothesis 4: Effort will depend solely to the piece rate the employer offered to the agent.

Hypothesis 5: The fixed wage will have no influence on the effort level chosen by the agent.

Hypothesis 6: The noise of the performance measure will have no influence on the effort level chosen by the agent.

Lastly, according to the model the principal is assumed to know the coefficient of absolute risk aversion of the agent in order to determine the optimal piece rate Although, we had initially examined the possibility of obtaining the coefficient of risk aversion of the agents in advance and provide it to the principals we believe that this information would be very difficult to be interpreted from the subjects. Therefore, we preferred to rely on the concept of social projection (Orbell and Dawes 1991). According to social projection, each player will project his own characteristics to others and use them as a cue on how they are more likely to behave.[17] If we assume that the principal will use social projection to infer how risk averse the agent they are matched with is, then we can formulate the following hypothesis.

Hypothesis 7: The more risk averse the principal is, the more risk averse he or she will expect the agent to be, and as a result the smaller the piece rate that he or she will offer to the agent.

5. Results

In this section we firstly present the descriptive statistics and the results of non-parametric tests in differences in piece rates, fixed wages and effort levels across treatments. All tests reported here are two tailed tests and have been

[17] For a review of the literature in psychology on social projection see Krueger (2007).

conducted at session level to control for non independence of observations. We continue with regression analysis with respect to the piece rates and effort.

4.1 *Non-Parametric Tests*

Figures 1 to 3 presents the average fixed wage, piece rate and effort level across treatments respectively. Table 1 summarises the average fixed wage, piece rate and effort across treatments. Qualitatively, when there is uncertainty in the environment (R treatment) employers decide to offer on average a higher fixed wage and a lower piece rate than when there is no uncertainty (NR). In addition, the effort level is smaller in the R treatment than in the NR treatment and almost the same as in the BR treatment. We conducted Mann-Whitney tests for fixed wages, piece rates and effort levels across all three treatments. Table 2 provides a summary of these tests. As can be seen in Table 2, only the difference between piece rates in in the R and NR treatments are statistically signinficant (Mann-Whitney p = 0.035).

Table 1: Average Fixed Wages, Piece Rates, and Efforts across Treatments

	Fixed wage	**Piece rate**	**Effort**
NR	69.93	16.51	7.44
R	72.75	14.41	6.94
BR	78.94	15.38	7.00

Result 1: In line with hypothesis 1, the piece rate is significantly smaller in the risk treatment than in the NR treatment.

One puzzling result worth investigating in future research has been that the piece rate in the NR treatment has been significantly smaller than the optimal and profit maximising piece rate of 20 (Mann-Whitney p < 0.001).

Result 2: In contrast to hypothesis 2, on average the piece rate offered in the NR treatment was statistically significantly smaller than the optimal piece rate of 20.

Table 2: Mann-Whitney tests for the fixed wage, piece rate and effort accross treatments

	Fixed wage	Piece rate	Effort
NR - R	0.762	0.034**	0.104
R - BR	0.273	0.199	0.734
NR -BR	0.131	0.325	0.161

Notes: All tests reported here are two tailed tests and have been conducted at session level to control for non independence of observations.

Interestingly the average piece rate that was offered by the principals in the R treatment was significantly larger than the piece rate predicted by the model (Mann-Whitney p = 0.023).

Result 3: In contrast to hypothesis 3, on average the piece rate offered in the R treatment was statistically significantly larger than the predicted piece rate of 13.35.

By conducting Spearman correlation tests between the three variables (piece rates, fixed wages, and effort levels), we found a very strong positive correlation between effort and the piece rate (as expected) of 0.81. In addition, we find a negative correlation between the fixed wage and the piece rate of -0.454 which suggest the employers would use the two mechanisms as substitutes to each other.

Figure 9: Average Fixed Wage across Treatments

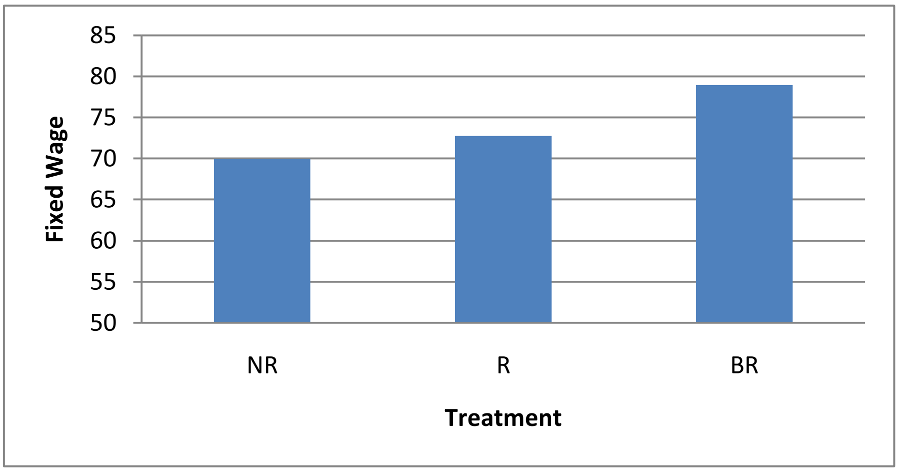


Figure 2: Average Piece Rate across Treatments

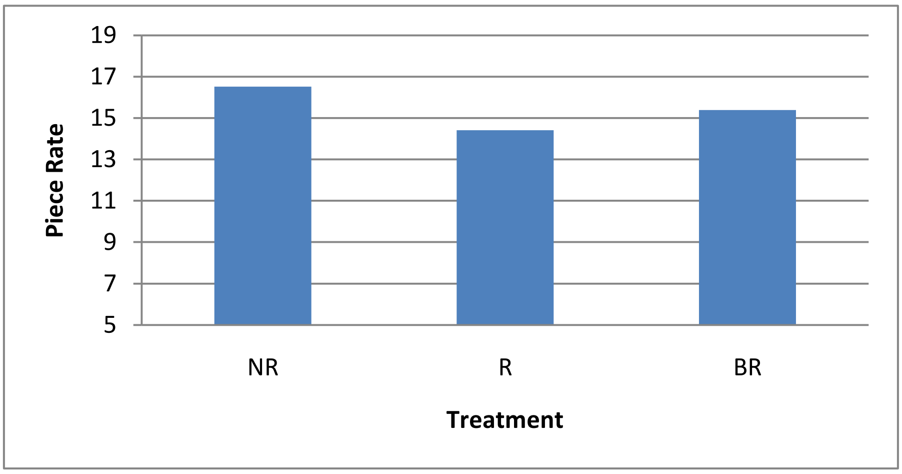
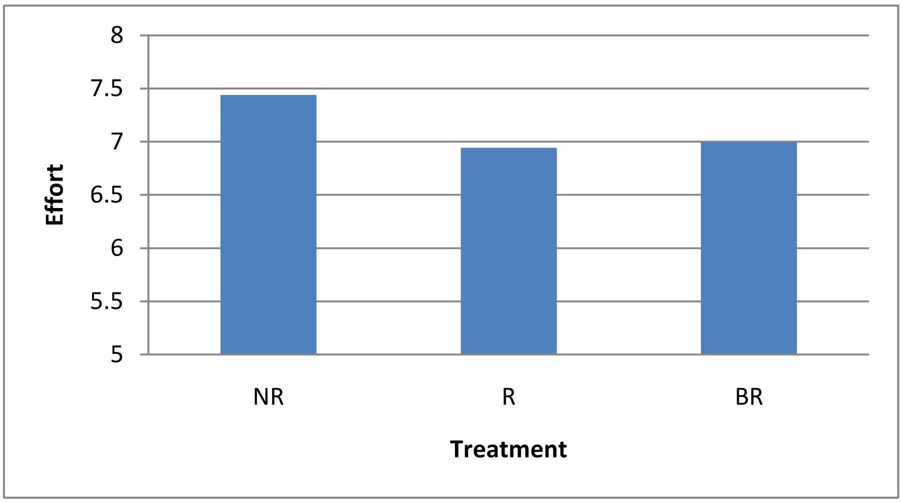


Figure 3: Average Effort across Treatments



4.2. *Regression Analysis*

In this section we report the results of panel regressions with random effects at subject level and error-clustering at session level firstly for the piece rates (Table 3) and secondly for effort (Table 4).[18] We retained one observation per round for

[18] We also conducted hierarchical linear models both for the piece rate and effort using random effects at subject level and session level with qualitatively similar results.

each subject leading to 1260 observations.[19] The regressions employ dummy variables for the experimental treatments. The NR treatment is used as a baseline. As three different measures of risk, (Holt and Laury, risk taking, and the risk perception) were collected, in each regression only one of them is used at each time. All three different measure of risk has been centered.[20] In addition, in the regressions in Table 3 although in theory we should have used the Holt and Laury scores of the employees assuming that the employers were able to guess the level of risk aversion of the employee they were matched with, we considered such an assumption implausible and we decided to use the *r* coefficient of the employer assuming the employer would expect the employee to be as risk averse (or risk loving) as he or she is (relying on the literature on social projection). In addition, we used interaction variables between each of the risk elicitation measure and the R treatment to capture any potential interaction effects between risk attitudes and the R treatment. Lastly, we used dummy variables for British students, gender, and economic students.

The results from the regressions on the piece rate in Table 3 are in parallel with the findings of the non parametric tests. In particular, in all the regressions that are presented in Table 1 the dummy variable for the R treatment is statistically significant at the 5% level reinstating Result 1 that the higher the variance in the environment the smaller the piece rate offered by the employers. In contrast, our control treatment (BR) was not statistically significant suggesting that the artificiality in which we imposed risk neutrality to the agent worked as we hypothesised and our result is robust.

Result 3: There is no statistically significant difference between the average piece rate offered in the BR treatment and the NR treament.

In regressions 2 to 4 we find that risk aversion has no impact on the size of the piece rate. However, after we introduced interaction variables between the risk treatment and each of the risk elicitation measures we observe a statistically

[19] In one of the sessions due to technical issues the choices of the subjects on the psychology questionnaires were not recorded, as a consequence in the regressions which employ the psychology questionnaires as a dependant variable there are 1218 observations as that session is omitted.

[20] See Dalal and Zickar (2012) for a recent discussion on the advantages of centering.

significant impact of the risk aversion coefficients both for the Holt and Laury and the risk taking tasks.

Result 4: In line with hypothesis 5, assuming social projection, we find some evidence that the more risk averse the employer believes the employee is, the smaller the piece rate that is observed.

Turning our attention to the regressions on the effort levels in Table 4, we observe that the treatment dummies have no signinficant effect in the effort levels as suggested by the theory. This result is persistent across all treatments. This finding is in line with the model's prediction and the findings of Sloof and van Praag (2008).

Result 5: In line with hypothesis 4, the effort choices of the employees are not affected by the variance in the performance measure.

However, we find highly statistically signifficant positive coefficients for both the fixed wage and the piece rates. Finding a positive statistically signifficant coefficient for the piece rate is in line with the model's prediction and profit maximising behaviour assumed in standard economic theory. However the significance of the fixed wage it could only be explained assuming social prefferences and/or reciprocal behaviour.

Result 6: In line with hypothesis 2, the employees responded with higher effort the higher the piece rate that was offered by the employers.

Result 7: In contrast to hypothesis 3, the employees responded with higher effort the higher the fixed wage that was offered by the employers.

Table 3: Panel regressions on piece rate (β) with random effects at subject level and error clustering at session level

	Regression 1	**Regression 2**	**Regression 3**	**Regression 4**	**Regression 5**	**Regression 6**	**Regression 7**
Piece Rate (β)							
R	-1.623**	-1.548**	-1.780**	-1703**	-1.518**	-1.875**	-1.715**
	(0.72)	(0.74)	(0.70)	(0.76)	(0.71)	(0.65)	(0.69)
BR	-0.203	-0.157	-0.461	-0.376	-0.119	-0.587	-0.383
	(0.69)	(0.66)	(0.64)	(0.72)	(0.65)	(0.70)	(0.70)
Fixed Wage	-0.053****	-0.053****	-0.050****	-0.050****	-0.053****	-0.050****	-0.050****
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Suggested Effort	1.198****	1.199****	1.126****	1.122****	1.198****	1.119****	1.121****
	(0.22)	(0.22)	(0.21)	(0.21)	(0.22)	(0.21)	(0.21)
Holt and Laury (c)		-0.172			-0.345**		
		(0.16)			(0.17)		
Risk Taking (c)			-0.055			-0.116**	
			(0.05)			(0.05)	
Risk Perception (c)				-0.043			-0.046
				(0.04)			(0.04)
R **x** Holt and Laury					0.558*		
					(0.32)		
R **x** Risk Taking						0.179*	
						(0.09)	
R **x** Risk Perception							0.010
							(0.10)
British	1.416*	1.359*	1.512*	1.573*	1.277*	1.537*	1.573*
	(0.78)	(0.77)	(0.80)	(0.83)	(0.77)	(0.80)	(0.84)
Gender	-0.176	-0.138	-0.294	-0.223	-0.124	-0.281	-0.231
	(0.64)	(0.64)	(0.64)	(0.64)	(0.64)	(0.66)	(0.62)
Economics Students	-1.093	-1.089	-1.152	-0.980	-0.956	-1.402	-0.973
	(1.47)	(1.42)	(1.47)	(1.46)	(1.42)	(1.40)	(1.44)
Constant	8.163****	8.104****	8.781****	8.734****	8.075****	8.943****	8.750****
	(1.97)	(1.42)	(1.84)	(1.87)	(1.86)	(1.81)	(1.84)
Obs	1260	1260	1218	1218	1260	1218	1218
R^2	0.24	0.24	0.24	0.23	0.25	0.25	0.23
Prob > X^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: * p < 0.1, ** p<0.05,*** p<0.01,**** p<0.001, standard errors provided in parentheses. In regressions where the Risk Taking and Risk Perception measures are employed (Reg. 3, 4,6, and 7) there are 42 less observations as in one of the sessions these variables were not recorded.

Table 4: Panel regressions on effort (*e*) with random effects at subject level and error clustering at session level

	Regression 1	**Regression 2**	**Regression 3**	**Regression 4**	**Regression 5**	**Regression 6**	**Regression 7**
Effort (*e*)							
R	0.077	0.094	0.123	0.093	0.086	0.115	0.044
	(0.14)	(0.14)	(0.14)	(0.15)	(0.15)	(0.15)	(0.14)
BR	-0.232	-0.209	-0.127	-0.216	-0.225	-0.139	-0.217
	(0.21)	(0.21)	(0.22)	(0.22)	(0.21)	(0.22)	(0.22)
Fixed Wage	0.012****	0.012****	0.012****	0.012****	0.012****	0.012****	0.012****
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Piece Rate	0.285****	0.285****	0.286****	0.285****	0.285****	0.285****	0.286****
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Suggested Effort	-0.018	-0.017	-0.016	-0.016	-0.016	-0.016	-0.018
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Holt and Laury (c)		-0.020			-0.004		
		(0.03)			(0.04)		
Risk Taking (c)			-0.020			- 0.017	
			(0.01)			(0.02)	
Risk Perception (c)				-0.004			-0.016
				(0.01)			(0.01)
R **x** Holt and Laury					-0.055		
					(0.93)		
R **x** Risk Taking						0.011	
						(0.03)	
R **x** Risk Perception							0.040
							(0.03)
British	0.252	0.242	0.187	0.233	0.236	0.188	0.259
	(0.17)	(0.17)	(0.18)	(0.18)	(0.17)	(0.18)	(0.17)
Gender	0.184	0.191	0.220	0.205	0.202	0.230	0.236
	(0.17)	(0.17)	(0.16)	(0.18)	(0.17)	(0.16)	(0.18)
Economics Students	-0.389	-0.369	-0.357	-0.400	-0.365	-0.364	-0.317
	(0.34)	(0.26)	(0.26)	(0.26)	(0.27)	(0.26)	(0.30)
Constant	1.902****	1.879****	1.818****	1.863****	1.885****	1.885****	1.842****
	(0.34)	(0.35)	(0.35)	(0.36)	(0.35)	(0.35)	(0.36)
Obs	1260	1260	1218	1218	1218	1218	1218
R^2	0.49	0.49	0.49	0.48	0.48	0.49	0.48
Prob > X^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: * p < 0.1, ** p<0.05,*** p<0.01,**** p<0.001, standard errors provided in parentheses. In regressions where the Risk Taking and Risk Perception measures are employed (Reg. 3, 4,6, and 7) there are 42 less observations as in one of the sessions these variables were not recorded.

6. Discussion

Our findings with respect to the relationship between risk and incentive intensity are in line with the predictions of the Incentive Intensity Principle. As a result our evidence provide support to the argument that previous empirical studies which found a positive or no relationship at all may have been due to the variety of other factors which could be affecting their data and are not included in the initial model. As a result our findings provide support to theorists who introduced alternative dimensions or additional variables in order to explain the observed positive relationship in previous studies (e.g. Prendergast, 2002).

A puzzling result has been that in the treatment where there was no variance on the output (NR treatment) the observed piece rates were significantly smaller than the optimal piece rate. This finding is in contrast to similar studies which observed that the employers not only offered contracts which are incentive compatible (Anderhub, et al. 2002, but also the first chapter of this thesis) but also which were more generous than the theoretical prediction. One explanation could be that the employers were expecting the employee to exert a higher effort due to the fixed wage that was also offered along with the piece rate. An alternative explanation may be that the use of a minimum wage, (which was introduced to ensure no loses) may created a reference point leading to the employers to offer smaller wages.[21] However, previous studies which focused on the effects of minimum wages have found that the use of minimum wages increases the average wages offered (Owen and Kagel, 2010) and the reservation utilities of the employees (Falk et al. 2006). Yet in these studies the experimental games resembled the gift exchange which is not possible to do an incentive compatible contract offer (except of a 0 offer) and necessitates pro-social behaviour.

Interestingly we find that the employees responded with higher effort levels for both higher piece rates and fixed wages. The response with higher effort the higher the piece rate offered (ceteris paribus) is in line with the predictions of the model, and more generally with the assumption of profit maximising behaviour. However, finding a positive relationship between the fixed wage and effort can only be explained with social preferences. Indeed, a positive relationship between the fixed wage and effort is in line with previous experimental

[21] A relevant and related literature to this interpretation would be the one on demand effects. For a discussion on demand effects see Zizzo (2010).

studies (Falk et al. 1999; Fehr et. al, 1998; Fehr and Gachter, 2002; Fehr and Schmidt, 2004, 2007).

It could be argued that a limitation of this study is that effort is measured in an abstract manner by allocating a decision number (i.e. subjects had to choose an effort level ranging from 5 to 10). This may create external validity concerns as compared to subjects performing a real effort task (Sloof and van Praag, 2008). However, using real effort tasks would generate different marginal costs of effort for each agent as subjects would differ in ability. This in turn creates an extra layer of complexity for the employer who would have not known what the cost of effort faced by the agent is. Therefore, although a real effort task could increase external validity, in order to minimise noise we chose to use an artificial effort task.

7. Conclusion

To conclude, despite previous studies have found mixed evidence regarding the relationship between risk and incentive intensity, our findings are in line with the negative relationship expected by the model. Our findings are the first to test experimentally the trade-off between risk and incentives which allowed for greater control and formulating an environment as close as possible to the model. In addition, we find no relation between the variance in the performance and the effort choice of the agent as well as a strong positive relation between the effort choice of the agents and the piece rate offered by the principal, also in line with the model's prediction. Moreover, the agents seemed to respond positively to higher fixed wages suggesting reciprocal behaviour which is in line with previous experimental studies on labour contracts (and particular the literature on gift exchange. However, we observed that the majority of the offers in the no risk treatment were offering a suboptimal piece rate which is surprisingly different to previous studies. A potential explanation for the lower offers may be our use of a minimum wage which may have acted as a reference point for the employers driving downwards the offers of the principals. Future research could investigate if the introduction of a low minimum wage leads to employment suboptimal contract offers.

References

Aggarwal, R. & Samwick, A. 1998. “The Other Side of the Trade-off: The Impact of Risk on Executive Compensation.” *Journal of Political Economy*, 107, 65-105.

Anderhub, V., Gächter, S., and Königstein, M. 2002. “Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment.”, *Experimental Economics*, Vol. 5, pp. 5–27.

Blais, A. and Weber, E. 2006. “Domain-Specific Risk-Taking (DOSPERT) scale for adult populations.” *Judgement and Decision Making*, 1, pp. 33-47.

Brown, C. 1990. "Firms Choice of Method of Pay." *Industrial and Labour Relations Review*, 43, pp. 165-82.

Bushman, R., Indejikian, R. and Smith, A. 1996. "CEO Compensation: The Role of Individual Performance Evaluation." *Journal of Accounting Economics*, 21, pp. 161-193.

Cooper, D. J., Kagel, J. H., 2003. “The impact of meaningful context on strategic play in signaling games.” *Journal of Economic Behavior and Organization*, 50, pp. 311-337.

Cooper, D. J., Kagel, J. H., 2009. “The role of context and team play in cross-game learning.” *Journal of the European Economic Association*, 7, pp. 1101-1139.

Dalal, D. and Zickar, M. (2012) “Some Common Myths About Centering Predictor Variables in Moderated Multiple Regression and Polynomial Regression” *Organizational Research Methods*, 15 pp. 339-362.

Eckel, C. C., & Grossman, P. J. 2002. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23, 281–295.

Eckel, C. C., & Grossman, P. J. 2008. “Forecasting risk attitudes: An experimental study of actual and forecast risk attitudes of women and men.” *Journal of Economic Behavior & Organization*, 68, pp. 1-17.

Englmaier, F. and Wambach, A. 2010. “Optimal Incentive Contracts under Inequity Aversion” *Games and Economic Behaviour*, 69, pp. 312-328.

Falk, A., Gächter, S. & Kovács, J., 1999. Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *Journal of Economic Psychology*, 20(3), pp.251–284.

Fehr, E. & Schmidt, K.M., 2007. Adding a Stick to the Carrot? The Interaction of Bonuses and Fines. *The American Economic Review*, 97, pp.177–181.

Fehr, E. and Schmidt, K.M., 2004. Fairness and Incentives in a Multi-Task Principal-Agent Model. *Scandinavian Journal of Economics*, 106, pp. 453-474.

Fehr, E. and Gächter, S., 2002. "Do Incentive Contracts Undermine Voluntary Cooperation?" Institute for Empirical Research in Economics, IEW Working Paper 34, University of Zurich.

Fehr, E., Kirchsteiger, G. & Riedl, A., 1998. Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, 2921(42), pp.1–34.

Fehr, E., Klein, A. and Schmidt, K.M., 2007. Fairness and Contract Design. *Econometrica*, 75(1), pp.121–154.

Fehr, E., and Schmidt, K.M., 1999. "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, 114, pp.817–868.

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

Garen, F. 1994. Executive Compensation and Principal-Agent Theory." *Journal of Political Economy*, 102, pp. 1175-1199.

Harrison, G. and Rutström E. 2008. "Risk Aversion in the Laboratory." In Jim C. Cox & Glenn W. Harrison (eds.), *Risk Aversion in Experiments, Research in Experimental Economics*, 12, Bingley, UK.

Hey, J. D., and Orme, C. 1994. "Investigating generalizations of expected utility theory using experimental data." *Econometrica*, 62, pp. 1291–1326.

Holmstrom, B. and Milgrom, R. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives" *Econometrica*, 55, 303-328.

Holt, C.A. 1995. Industrial organization: A survey of laboratory research. In: Kakel, J.H. and Roth, A.E. (Eds.), *The Handbook of Experimental Economics*. Princeton University Press, Princeton, 349-443.

Holt, C., and Laury, S. 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92, 1644-1655.

Krueger, J. 2007. "From social projection to social behaviour." *European Review of Social Psychology*, 18, pp. 1-35

Lafontaine, F. 1992. "Agency Theory and Franchising: Some empirical results." *The RAND Journal of Economics,* 23, 263-283.

Norton C. 1988. "An Empirical Look At Franchising As An Organizational Form", *Journal of Business*, 61, pp. 197-218.

Prendergast, C. 1999. "The Provision of Incentives in Firms" *Journal of Economic Literature*, 37, 7-63.

Prendergast, C. 2000. "What Trade-off of Risk and Incentives." *American Economic Review AEA Papers and Proceedings,* 90, 421-425.

Prendergast, C. 2002. "The Tenuous Trade-off Between Risk and Incentives" *Journal of Political Economy*, 110, 1071-1102.

Sloof, R. and van Praag, M. 2008. "Performance Measurement, Expectancy and Agency Theory: An Experimental Study." *Journal of Economic Behavior and Organization*, 67, 794-809.

Zizzo, D.J. 2010. "Experimenter Demand Effects in Economic Experiments." *Experimental Economics*. 13, 75-98.

Appendix A - Instructions

Instructions (NR Treatment)

Welcome to our experiment! You are participating in an experiment on decision making. The experiment is expected to last no more than 1 hour and 15 minutes. Please read the following instructions carefully. During the experiment, you are not allowed to communicate with other participants. If you face any questions at any moment please raise your hand and the experimenter will come to your desk.

In the experiment you will be using an experimental currency called ECU. In the end of the experiment the ECU you have earned during the experiment will be converted at the exchange rate of: **1ECU = £0.02.**

For example, 10ECU=£0.20, 100ECU=£2, 50ECU= £1, 200ECU= £4.

Experiment Overview

The experiment consists of two parts. The first part is explained in detail bellow the second consists of two questionnaires and will be explained at the end of the first part.

The first part of the experiment consists of **10 rounds.** The **first 3 rounds** are **practice** rounds, this means that your choices will not affect your earnings. Their role is to help you understand and familiarise with the tasks involved. One from the following 7 rounds will be chosen randomly by the computer and paid to you in cash at the end of the experiment.

Each participant is assigned randomly the role of either the **employer** or the **employee**. You hold this role throughout the experiment. If you are an employer the computer will **randomly match** you with an employee at the start of every round and if you are an employee with an employer. The experiment is **anonymous**; this means that you will not know with whom of the other participants you are interacting.

The Structure of a Round

1) The employer has to offer an employment contract to the employee requesting him/her to exert a level of effort.

2) The employee decides to accept or reject the contract:
    a. If he/she rejects the contract the round finishes and both earn 200 ECU.
    b. If the employee accepts the contract both receive the 200 ECU and he/she decides what effort level he/she wants to exert.

3) After the employee has chosen an effort level the computer calculates the profits of both and the round finishes.

The Contract

If you are an **employer** you need to decide what effort level you want the employee to exert. After you decide the effort you would want the employee to exert you need to think what contract to offer to the employee given the implications that this has in earnings. If you are the **employee** what you need to think is what effort you would want to exert for the given contract taking into account the effect this has on earnings.

At the start of every round both employer and employee receive **200 ECU.** This money is for you to use within the experiment and are added to your profits for the round.

Effort in this experiment is represented by a number the employee chooses which ranges from 5 to 10. **Every unit of effort costs ECU to the employee**. Table 1 shows the corresponding employer revenue and cost of effort for each unit of effort.

The revenue of the employer is determined by the following:

Revenue of the employer: 50 x effort.

That means that for every unit of effort the employee exerts the employer earns 50 ECU. For example, if the employee exerts an effort of 3 the employer earns 150 ECU.

Table 1: Effort levels, Cost of Effort, and Employer Revenue

Effort	Cost of effort	Employer Revenue
5	25	250
6	36	300
7	49	350
8	64	400
9	81	450
10	100	500

Employers can choose to pay the employee with a **fixed wage** and/or with a **piece rate**.

A **fixed wage** is a transfer of money from the employer to the employee which is independent of how much effort he/she exerts (for example, a salary). The fixed wage can **range from 0 to 200 ECU**.

A **piece rate** is a payment for **every unit of effort.** An example of that could be an apple picker. If the employee was an apple picker a piece rate would mean a specific amount of money (ECU) for every basket of apples (effort) he brings to the employer. For example for a piece rate of 20, and an effort level of 5 the employee will be paid 20x5= 100 ECU i.e. the employee earns 100 ECU. (**The piece rate can range from 0 to 40** including one decimal (i.e. 10.1, 23.4, 30.5 etc.).

Suggested effort

In his/her contract offer the employer has to suggest an effort level to the employee. Note however that the **suggested effort** of the employer is only a suggestion. The employee is not bound to that suggestion but he/she is free to choose any effort level within the given range of 0 to 5.

The Minimum Contract

The minimum effort of the employee is **5**. To ensure that the employee is at least compensated for his/her minimum effort the contract offered by the employer **must** have a minimum fixed wage of 50 ECU and a piece rate of 5.

How Earnings from a Round are calculated

For the employer his/her earnings are the 200 ECU he received at the start of the round, **plus** the revenue generated by the employee's effort, **minus** the fixed wage he/she paid and **minus** the piece rate he/she paid. In other words:

Employer's Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate **x** Effort

In the case of the employee, his/her earnings are his/her 100 ECU **plus** the fixed wage **plus** the piece rate **times** the effort, **minus** the cost of effort. In other words:

Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate **x** Effort – Cost of Effort

Overview

5. The employer chooses the **fixed wage** (from 0 to 350), **the piece rate** (0 to 50) and **suggests an effort** level (from 5 to 10) to the employee.
6. Afterwards, the employee has been informed of the offered contract, he/she has to decide either to **accept or reject** the contract. If the employee rejects the contract the stage finishes. If he accepts the contract, he receives the offered fixed wage and **decides an effort level** (from 5 to 10).
7. Once the employee has decided an effort level, **the computer calculates the earnings** of the employer and the employee and informs both participants.
8. This procedure is repeated till we reach round 10.

Some Examples

Think the following examples carefully and try to see if the earnings have been calculated correctly. The numbers chosen are purely illustrative.

Example 1:

Employer Choices	
Fixed Wage	200
Piece Rate	0
Suggested Effort	10

Employee Choice	
Effort	10

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	450
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	250

Example 2:

Employer Choices	
Fixed Wage	350
Piece Rate	0
Suggested Effort	10

Employee Choice	
Effort	5

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	50
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	475

Example 3:

Employer Choices	
Fixed Wage	25
Piece Rate	20
Suggested Effort	10

Employee Choice	
Effort	10

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	425
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	275

Example 4:

Employer Choices	
Fixed Wage	25
Piece Rate	30
Suggested Effort	8

Employee Choice	
Effort	10

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	325
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	375

Example 5:

Employer Choices	
Fixed Wage	50
Piece Rate	24.7
Suggested Effort	5

Employee Choice	
Effort	5

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	353
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	347

Instructions (R Treatment)

Welcome to our experiment! You are participating in an experiment on decision making. The experiment is expected to last no more than 1 hour and 15 minutes. Please read the following instructions carefully. During the experiment, you are not allowed to communicate with other participants. If you face any questions at any moment please raise your hand and the experimenter will come to your desk.

In the experiment you will be using an experimental currency called ECU. In the end of the experiment the ECU you have earned during the experiment will be converted at the exchange rate of: **1ECU = £0.02.**

For example, 10ECU=£0.20, 100ECU=£2, 50ECU= £1, 200ECU= £4.

Experiment Overview

The experiment consists of two parts. The first part is explained in detail bellow the second consists of two questionnaires and will be explained at the end of the first part.

The first part of the experiment consists of **10 rounds.** The **first 3 rounds** are **practice** rounds, this means that your choices will not affect your earnings. Their role is to help you understand and familiarise with the tasks involved. One from the following 7 rounds will be chosen randomly by the computer and paid to you in cash at the end of the experiment.

Each participant is assigned randomly the role of either the **employer** or the **employee**. You hold this role throughout the experiment. If you are an employer the computer will **randomly match** you with an employee at the start of every round and if you are an employee with an employer. The experiment is **anonymous**; this means that you will not know with whom of the other participants you are interacting.

The Structure of a Round

4) The employer has to offer an employment contract to the employee requesting him/her to exert a level of effort.

5) The employee decides to accept or reject the contract:

   a. If he/she rejects the contract the round finishes and both earn 200 ECU.
   b. If the employee accepts the contract both receive the 200 ECU and he/she decides what effort level he/she wants to exert.

6) After the employee has chosen an effort level the computer calculates the profits of both and the round finishes.

The Contract

If you are an **employer** you need to decide what effort level you want the employee to exert. After you decide the effort you would want the employee to exert you need to think what contract to offer to the employee given the implications that this has in earnings. If you are the **employee** what you need to think is what effort you would want to exert for the given contract taking into account the effect this has on earnings.

At the start of every round both employer and employee receive **200 ECU.** This money is for you to use within the experiment and are added to your profits for the round.

Effort in this experiment is represented by a number the employee chooses which ranges from 5 to 10. **Every unit of effort costs ECU to the employee**. Table 1 shows the corresponding employer revenue and cost of effort for each unit of effort.

The revenue of the employer is determined by the following:

Revenue of the employer: 50 x effort.

That means that for every unit of effort the employee exerts the employer earns 50 ECU. For example, if the employee exerts an effort of 3 the employer earns 150 ECU.

Table 1: Effort levels, Cost of Effort, and Employer Revenue

Effort	Cost of effort	Employer Revenue
5	25	250
6	36	300
7	49	350
8	64	400
9	81	450
10	100	500

Employers can choose to pay the employee with a **fixed wage** and/or with a **piece rate**.

A **fixed wage** is a transfer of money from the employer to the employee which is independent of how much effort he/she exerts (for example, a salary). The fixed wage can **range from 0 to 200 ECU**.

A **piece rate** is a payment for **every unit of output. Output is the sum of effort + a luck value**. An example of that could be an apple picker. If the employee was an apple picker a piece rate would mean a specific amount of money (ECU) for every basket of apples (output) he brings to the employer. And the luck factor could be how favourable or unfavourable the weather conditions has been. The piece rate can **range from 5 to 40**. For example, for a piece rate of 10, an effort level of 3, and a luck value of 2, it means the employee will be paid 10x(3+2)= 50 i.e. the employee earns 50 ECU.

Table 2 show the values luck may take (that is from -5 to 5) and what is the chance for each of these values to be selected by the computer. For example, the chance the luck value to turn out to be -5 is one out of a hundred, for -1 is sixteen out of a hundred, for 2 is twelve out of a hundred etc.

Table 2: Chance of each luck factor to happen

Luck	Chance
-5	1%
-4	4%
-3	7%
-2	12%
-1	16%
0	18%
1	16%
2	12%
3	7%
4	4%
5	1%

The earnings for the employee from the piece rate are calculated by the ECU value chosen from the employer (from 5 to 40) multiplied by the output which is the sum of the effort and luck (i.e. piece rate **x** (effort + luck)). Note that luck only affects the earnings of the employee.

An example (numbers are purely illustrative)

Assume: a) the employer chooses to set the piece rate at the value of 5, b) the employee chooses an effort of 5 and the luck value turns out to be -2.

The employee's from the piece rate are the piece rate times the sum of effort and luck (i.e. 5 x (5+(-2))= 5 x 3 = 15) which is 15 ECU plus the 100 ECU minus the cost of effort for 5 units of effort which is 28. Therefore the employee will earn 115 – 28 which is 87 ECU.

For the employer however the earnings are calculated without considering the luck value. Therefore, the employer will receive 25 x 5 (the revenue from 5 units of effort) minus the piece rate which will be 5 (the piece rate) times the effort of the employee which was 5 plus the 100 ECU that is given to him at the start of the round. Hence 125 – 25 + 100 leading to earnings of 100 ECU.

Suggested effort

In his/her contract offer the employer has to suggest an effort level to the employee. Note however that the **suggested effort** of the employer is **only a suggestion**. The employee is not bound to that suggestion but he/she is free to choose any effort level within the given range of 0 to 5.

The Minimum Contract

The minimum effort of the employee is **5**. To ensure that the employee is at least compensated for his/her minimum effort the contract offered by the employer **must** have a minimum fixed wage of 50 ECU and a piece rate of 5.

How Earnings from a Round are calculated

For the employer his/her earnings are the 200 ECU he received at the start of the round, **plus** the revenue generated by the employee's effort, **minus** the fixed wage he/she paid and **minus** the piece rate he/she paid. In other words:

Employer's Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort

In the case of the employee, his/her earnings are his/her 100 ECU **plus** the fixed wage **plus** the piece rate **times the sum** of effort and the luck factor, **minus** the cost of effort. In other words:

Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate **x** (Effort+ luck) – Cost of Effort

Note that the luck factor affects only the employee!

Overview

9. The employer chooses the **fixed wage** (from 50 to 200), **the piece rate** (5 to 40) and **suggests an effort** level (from 5 to 10) to the employee.
10. Afterwards, the employee has been informed of the offered contract, he/she has to decide either to **accept or reject** the contract. If the employee rejects the contract the stage finishes. If he accepts the contract, he receives the offered fixed wage and **decides an effort level** (from 5 to 10).
11. Once the employee has decided an effort level, **the computer calculates the earnings** of the employer and the employee and informs both participants.
12. This procedure is repeated till we reach round 10.

Some Examples

Think the following examples carefully and try to see if the earnings have been calculated correctly. The numbers chosen are purely illustrative.

Example 1:

Employer Choices	
Fixed Wage	200
Piece Rate	5
Suggested Effort	10

Employee Choice	
Effort	10

Computer Choices	
Luck	0

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	450
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	350

Example 2:

Employer Choices	
Fixed Wage	200
Piece Rate	10
Suggested Effort	10

Employee Choice	
Effort	5

Computer Choices	
Luck	-2

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	200
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	405

Example 3:

Employer Choices	
Fixed Wage	50
Piece Rate	20
Suggested Effort	10

Employee Choice	
Effort	10

Computer Choices	
Luck	+2

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	450
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	390

Example 4:

Employer Choices	
Fixed Wage	50
Piece Rate	30
Suggested Effort	8

Employee Choice	
Effort	10

Computer Choices	
Luck	-5

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	350
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	300

Example 5:

Employer Choices	
Fixed Wage	50
Piece Rate	24.7
Suggested Effort	5

Employee Choice	
Effort	5

Computer Choices	
Luck	+5

Employer Earnings = 200 ECU + Revenue – Fixed Wage – Piece Rate x Effort	277
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x Effort – Cost of Effort	472

Instructions (BR Treatment)

Welcome to our experiment! You are participating in an experiment on decision making. The experiment is expected to last no more than 1 hour and 15 minutes. Please read the following instructions carefully. During the experiment, you are not allowed to communicate with other participants. If you face any questions at any moment please raise your hand and the experimenter will come to your desk.

In the experiment you will be using an experimental currency called ECU. In the end of the experiment the ECU you have earned during the experiment will be converted at the exchange rate of: **1ECU = £0.02.**

For example, 10ECU=£0.20, 100ECU=£2, 50ECU= £1, 200ECU= £4.

Experiment Overview

The experiment consists of two parts. The first part is explained in detail bellow the second consists of two questionnaires and will be explained at the end of the first part.

The first part of the experiment consists of **10 rounds.** The **first 3 rounds** are **practice** rounds, this means that your choices will not affect your earnings. Their role is to help you understand and familiarise with the tasks involved. One from the following 7 rounds will be chosen randomly by the computer and paid to you in cash at the end of the experiment.

Each participant is assigned randomly the role of either the **employer** or the **employee**. You hold this role throughout the experiment. If you are an employer the computer will **randomly match** you with an employee at the start of every round and if you are an employee with an employer. The experiment is **anonymous**; this means that you will not know with whom of the other participants you are interacting.

The Structure of a Round

7) The employer has to offer an employment contract to the employee requesting him/her to exert a level of effort.

8) The employee decides to accept or reject the contract:
    a. If he/she rejects the contract the round finishes and both earn 200 ECU.
    b. If the employee accepts the contract both receive the 200 ECU and he/she decides what effort level he/she wants to exert.

9) After the employee has chosen an effort level the computer calculates the profits of both and the round finishes.

The Contract

If you are an **employer** you need to decide what effort level you want the employee to exert. After you decide the effort you would want the employee to exert you need to think what contract to offer to the employee given the implications that this has in earnings. If you are the **employee** what you need to think is what effort you would want to exert for the given contract taking into account the effect this has on earnings.

At the start of every round both employer and employee receive **200 ECU.** This money is for you to use within the experiment and are added to your profits for the round.

Effort in this experiment is represented by a number the employee chooses which ranges from 5 to 10. **Every unit of effort costs ECU to the employee**. Table 1 shows the corresponding employer revenue and cost of effort for each unit of effort.

The revenue of the employer is determined by the following:

Revenue of the employer: 50 x (Effort + Luck).

Luck is a number that is randomly chosen by the computer and can range from -5 to 5.

That means that for every unit of effort (assume luck is 0) the employee decides, the employer earns 50 ECU. For example, if the employee decides an effort of 3 the employer earns 150 ECU.

Table 1: Effort levels, Cost of Effort, and Employer Revenue

Effort	Cost of effort	Employer Revenue
5	25	250
6	36	300
7	49	350
8	64	400
9	81	450
10	100	500

Employers can choose to pay the employee with a **fixed wage** and/or with a **piece rate**.

A **fixed wage** is a transfer of money from the employer to the employee which is independent of how much effort he/she exerts (for example, a salary). The fixed wage can **range from 50 to 200 ECU**.

A **piece rate** is a payment for **every unit of output. Output is the sum of effort and the luck value**. An example of that could be an apple picker. If the employee was an apple picker a piece rate would mean a specific amount of money (ECU) for every basket of apples (output) he brings to the employer. And the luck factor could be how favourable or unfavourable the weather conditions have been. The piece rate can **range from 5 to 40**. For example, for a piece rate of 10, an effort level of 5, and a luck value of 2, it means the employee will be paid 10x(5+2)= 70 i.e. the employee earns 70 ECU.

Table 2 show the values **luck** may take (that is from -5 to 5) and what is the chance for each of these values to be selected by the computer. For example, the chance the luck value to turn out to be -5 is one out of a hundred, for -1 is sixteen out of a hundred, for 2 is twelve out of a hundred etc.

Table 2: Chance of each luck factor to happen

Luck	Chance
-5	1%
-4	4%
-3	7%
-2	12%
-1	16%
0	18%
1	16%
2	12%
3	7%
4	4%
5	1%

The earnings for the employee from the piece rate are calculated by the ECU value chosen from the employer (from 5 to 40) multiplied by the output which is the sum of the effort and luck (i.e. piece rate **x** (effort + luck)).

An example (numbers are purely illustrative)

Assume: a) the employer chooses to set the piece rate at the value of 5, b) the employee chooses an effort of 5 and the luck value turns out to be -2.

The employee’s earnings from the piece rate are the piece rate times the sum of effort and luck (i.e. 5 **x** (5+(-2))= 5 **x** 3 = 15) which is 15 ECU plus the 200 ECU minus the cost of effort for 5 units of effort which is 28. Therefore the employee will earn 215 – 28 which is 187 ECU.

The employer will receive 50 **x** (5-2), which equals to 150 ECU, minus the piece rate which will be 15 plus the 200 ECU that is given to him at the start of the round. Hence 150 – 15 + 200 leading to earnings of 3250 ECU.

Suggested effort

In his/her contract offer the employer has to suggest an effort level to the employee. Note however that the **suggested effort** of the employer is **only a suggestion**. The employee is not bound to that suggestion but he/she is free to choose any effort level within the given range of 5 to 10.

The Minimum Contract

The minimum effort of the employee is **5**. To ensure that the employee is compensated for his/her minimum effort the contract offered by the employer **must** have a minimum fixed wage of 50 ECU and a piece rate of 5. This way it is ensured that neither the employers nor the employees can make losses.

How Earnings from a Round are calculated

For the employer his/her earnings are the 200 ECU he received at the start of the round, **plus** the revenue generated by the sum of the employee's effort and the luck factor, **minus** the fixed wage he/she paid and **minus** the piece rate he/she paid. In other words:

Employer's Earnings = 200 ECU **+** 50**x**(Effort+Luck) **–** Fixed Wage **–** Piece Rate **x** (Effort+Luck)

In the case of the employee, his/her earnings are his/her 200 ECU **plus** the fixed wage **plus** the piece rate **times the sum** of effort and the luck factor, **minus** the cost of effort. In other words:

Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate **x** (Effort+ luck) – Cost of Effort

Overview

13. The employer chooses the **fixed wage** (from 50 to 200), **the piece rate** (5 to 40) and **suggests an effort** level (from 5 to 10) to the employee.

14. Afterwards, the employee has been informed of the offered contract, he/she has to decide either to **accept or reject** the contract. If the employee rejects the contract the stage finishes. If he accepts the contract, he receives the offered fixed wage and **decides an effort level** (from 5 to 10).

15. Once the employee has decided an effort level, **the computer calculates the earnings** of the employer and the employee and informs both participants.

16. This procedure is repeated till we reach round 10.

Some Examples

Think the following examples carefully and try to see if the earnings have been calculated correctly. The numbers chosen are purely illustrative.

Example 1:

Employer Choices	
Fixed Wage	200
Piece Rate	5
Suggested Effort	10

Employee Choice	
Effort	10

Computer Choices	
Luck	0

Employer's Earnings = 200ECU **+** 50**x**(Effort+Luck) **–** Fixed Wage **–** Piece Rate **x** (Effort+Luck)	450
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x (Effort+Luck) – Cost of Effort	350

Example 2:

Employer Choices	
Fixed Wage	200
Piece Rate	10
Suggested Effort	10

Employee Choice	
Effort	5

Computer Choices	
Luck	-2

Employer's Earnings = 200ECU **+** 50**x**(Effort+Luck) **–** Fixed Wage **–** Piece Rate **x** (Effort+Luck)	120
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x (Effort+Luck) – Cost of Effort	405

Example 3:

Employer Choices	
Fixed Wage	50
Piece Rate	20
Suggested Effort	10

Employee Choice	
Effort	10

Computer Choices	
Luck	+2

Employer's Earnings = 200ECU **+** 50**x**(Effort+Luck) **–** Fixed Wage **–** Piece Rate **x** (Effort+Luck)	510
Employee's Earnings = 200 ECU + Fixed Wage + Piece Rate x (Effort+Luck) – Cost of Effort	390

Example 4:

Employer Choices	
Fixed Wage	50
Piece Rate	30
Suggested Effort	8

Employee Choice	
Effort	10

Computer Choices	
Luck	-5

Employer’s Earnings = 200ECU **+** 50**x**(Effort+Luck) **–** Fixed Wage **–** Piece Rate **x** (Effort+Luck)	250
Employee’s Earnings = 200 ECU + Fixed Wage + Piece Rate x (Effort+Luck) – Cost of Effort	300

Example 5:

Employer Choices	
Fixed Wage	50
Piece Rate	24.7
Suggested Effort	5

Employee Choice	
Effort	5

Computer Choices	
Luck	+5

Employer’s Earnings = 200ECU **+** 50**x**(Effort+Luck) **–** Fixed Wage **–** Piece Rate **x** (Effort+Luck)	403
Employee’s Earnings = 200 ECU + Fixed Wage + Piece Rate x (Effort+Luck) – Cost of Effort	472

Appendix B – The P-A and Theoretic Predictions

The Parameters of Experimental Implementation

There are two individuals a principal *P* and an agent *A*. The principal wants to hire the Agent to exert effort e. For every unit of effort exerted by the agent the principal earns 50 ECU[22]. However, effort is costly for the agent and is given by C(e). The principal uses a linear incentive scheme to hire the agent. When there is no risk and effort is observable the principal's profit function is defined as:

$$P^P = R(e) - [F + \beta(e)] \qquad (1)$$

The agent's profit when there is no risk is defined as:

$$P^A = F + \beta(e) - C(e) \qquad (2)$$

The *Total Revenue* is given by:

$$R(e) = 50 * e \qquad (3)$$

The cost of effort is a strictly increasing and convex function in effort:

$$C(e) = e^2 \qquad (4)$$

With:

F ... Unconditional fixed wage

$F \in \{50,51,\dots,199,200\}$

β ... The piece rate paid to the agent for each unit of effort.

$\beta \in \{5, 6, \dots,39,40\}$

e ... Effort level revealed by the agent

$e \in \{5,6,\dots,9,10\}$

[22] ECU stands for Experimental Currency Units

The game theoretic solution

Given the above parameters the *participation constraint*, i.e. the constraint that has to be met in order to make any contract offer monetarily beneficial is:

$$F + \beta(e) \geq C(e) \qquad (5)$$

Any offer that does not satisfy (5) if accepted would imply the agent would make losses.

The principal wants to:

$$\max(R(e) - F - \beta(e))$$

Given that the agent would accept any contract that satisfies (5) the minimum amount that has to be transferred to the agent has to be equal to $C(e)$. Hence,

$$F + \beta(e) = C(e)$$

The principal's maximization problem becomes

$$\max\,(R(e) - C(e)) \qquad (6)$$

Inserting the actual parameters used in the experiment results in

$$\max\,(50e - e^2)$$

Maximizing by e results in

$$e = 25$$

Thus the optimal effort level would be 25. As the experimental parameters only allow $e \in \{5,6,\ldots,9,10\}$, the maximisation problem in (6) has a *corner solution*[23] of $e^* = 20$.

Maximizing (2) with respect to e^* leads to

$$\beta - 2e^* = 0 \text{ and}$$

$$\beta = 2e^*$$

[23] As in the previous chapter , the choice for a corner solution was made to reduce complexity to an already highly complex design from the perspective of the principal. This choice though bears the cost that will be harder to test if the principals had correctly identified that e* is the optimum effort level or if they chose it ad hoc simply following a rule of thumb such as the more the better. Nevertheless, the use of corner solutions has been a standard approach to experiments which investigated contract design and social preferences.

Inserting the above calculated effort level $e^{*} = 10$ and solving for β, finally provides the minimum piece rate β^{*}. Thus,

$$\beta^{*} = 20 \tag{7}$$

Thus, the incentive contract is incentive compatible for any value of $\beta \geq 20$.

With $\beta = \beta^{*}$ the consequent profits (after taking into account their endowments) for the principal $P^{P} = 450$ and $P^{A} = 350$ for the agent.

When there is risk in the environment the principal is assumed to be able to observe the final output (i.e. Total Revenue) but he or she is unable to observe what part of this output is due to the agent's effort and what is due to randomness. In this case the total revenue function can be expressed the following way:

$$R(e) = 50 * (e + x) \tag{8}$$

With:

x ... Stochastic random factor

$x \sim N\ (0{,}2.5)$

Assuming the principal is risk neutral he or she will maximise expected profit:

$$E(P^{P}) = R(e) - (F + \beta(e + x)) \tag{9}$$

Given the principal is risk neutral and $x \sim N\ (0{,}2.5)$ the profit function of the principal can be re-written as:

$$E(P^{P}) = 50e - (F + \beta e) \tag{10}$$

However, given the agent is risk averse his or her expected profit is given by:

$$E(P^{A} = F + \beta e - C(e) - \frac{1}{2} r \beta^{2} var(x) \tag{11}$$

Where:

r = coefficient of absolute risk aversion of the agent

Given profit functions 10 and 11 Holmstrom and Milgrom, (1987) have shown that optimal incentive intensity β is given by:

$$\beta = \frac{R'(e)}{1 + rVC''(e)}$$

Where: $V = var(x)$

Proof

$P^A = F + \beta e - C(e) - \frac{1}{2} r\beta var(x)$

The agent will choose *e* to maximise his or her expected profit.

F.O.C.

$\beta = C'(e)$ (Incentive Constraint) (12)

Given the agent's reservation utility is assumed to be 0 the principal needs to satisfy the following participation constraint:

$F + \beta e = C(e) + \frac{1}{2} r\beta V$ (13)

Principal's net profit is given by:

$P^P = R(e) - (F + \beta e)$, substituting equation 13 leads to

$P^P = R(e) - C(e) - \frac{1}{2} r\beta V$, substituting equation 12 for β,

$P^P = R(e) - C(e) - \frac{1}{2} rC'(e)^2 V$

F.O.C.

$\frac{\partial P^P}{\partial e} = R'(e) - C'(e) - rVC'(e)C''(e) = 0 \Rightarrow$

$R'(e) = C'(e) + rVC'(e)C''(e) \Rightarrow$

$R'(e) = C'(e)[1 + rVC'(e)]$, substituting β for $C'(e)$

$R'(e) = \beta[1 + rVC'(e)]$ and solving for β

$\beta = \frac{R'(e)}{1 + rVC''(e)}$

Given that V in our experiment was set at 2.5 and the rest of the factors remained unchanged the optimal β^* is given by:

$$\beta^* = \frac{50}{1+5r}$$

We elicited the coefficient of absolute risk aversion for all subjects by using the Holt and Laury (2002) questionnaire. In order to generate a benchmark for our analysis we calculated an average $\bar{r}$ from all subjects in our experiment.[24] The average coefficient of absolute risk aversion from all subjects who participated in our experiment was $\bar{r}$ = 0.549. After inserting $\bar{r}$ in equation 1 it yields the optimal β^*= 13.35. Given the optimal β^* the optimal effort level for the employee is 7. The consequent expected profits (after taking into account their endowments) for the principal $P^P = 406$ and $P^A = 294$ for the agent.

[24] Note that the model assumes that the principal is aware of the exact value of the coefficient of absolute risk aversion for each agent.

Chapter 3: Obedience[25]

"I know the power obedience has of making things

easy which seem impossible"

Saint Teresa of Avila (1972:36)*, "The interior castle"*

1. Introduction

This chapter presents a simple experiment on the role of authority and obedience in an experiment where obedience damages other people's earnings. Principal agent models are the standard conceptual framework by which economists consider authority, but their focus is on analyzing the effect of different pay structures on decision making by agents. While social preferences have been usefully modeled within this framework (Englmeier and Wambach, 2010), consideration has not been given to how authority *per se* may help induce compliance. There has been considerable attention in economics to conformism and social norms, by which subjects tend to do what a number of others do (e.g. Asch, 1955; Jones, 1984; Lopez-Perez, 2008; Zafar, 2011), and there is of course a significant empirical literature on peer effects (e.g., Case and Katz, 1991; Kawaguchi, 2004; Powell et al., 2005; Lundborg, 2006) and on social image and pro-social behavior (e.g., Glazer and Konrad, 1996; Benabou and Tirole, 2006; Andreoni and Bernheim, 2009; Ariely et al., 2009), but the focus of this research has been on what we may label as *horizontal* social pressure, i.e. pressure by peers.[26] There has also been some insightful attention to the study of leadership as a

[25] This chapter is based on a paper co-authored with my supervisor Prof. Daniel Zizzo.

[26] The difficulties with identification in peer effects estimations imply that this is an area where experimental research is especially useful (e.g. Manski, 1993). In psychology, the connection is with

way of helping to solve social dilemmas or weakest link type coordination problems (Moxnes and van Heijden, 2003; Brandts et al., 2007; Guth et al., 2007; van der Heijden et al., 2008), but the leaders in this literature are just peers whose actions may facilitate cooperation and contribution within groups.

The question we wish to draw the attention of economists to with this chapter is instead the following: can *vertical*, i.e. hierarchical, social pressure induce conformism, even when there is no financial reason for obeying, and even when the domain for the action to be undertaken by the agent is anti-social, i.e. contrary to the standard social norm of not damaging others? If so, this should be taken into account in principal agent modeling and more generally in thinking about incentives and delegation in organizations. It is notable to contrast the economist's perspective, which does away with the motivation to obey an authority, with that of psychologists such as Cialdini and Goldstein (2004, p. 596) when they say that "most organizations would cease to operate efficiently if deference to authority were not one of the prevailing norms."

In this chapter we capture deference to authority as social image utility towards the authority. Intuitively, there is good reason to believe that subjects may typically care at least as much if not more of social image with respect to an authority than with respect to peers. Sliwka (2007) and Ellingsen and Johannesson (2008) contain important models that consider different aspects of how the social image that the agent feels as a response to the incentive structure by the principal matters, and may in turn shape the incentive contract chosen by the principal. Our starting point is the same, but our focus is on whether the agent has a drive to obey as a result of social image utility towards the authority. In this respect, the choice of an anti-social domain for the action to be undertaken by the agent, namely to destroy half of their partner's earnings, is especially useful to control for potential

traditional research on the value of conformism to peers (e.g., Asch, 1955) and on influence (e.g., Cialdini and Goldstein, 2004).

explanations of changes in behavior based on social norms to be pro-social (as in, e.g., Bicchieri, 2006; Keizer et al., 2008; Bicchieri and Xiao, 2009).

The focus on the domain of destructive behavior is of interest in its own right. No doubt, obedience as a character trait is generally perceived as a virtue that ought to be praised and promoted within society.[27] Equally, blind obedience can lead to catastrophic outcomes. In his defense on the court in Jerusalem for his role in the Holocaust and his crimes against the Jews, Eichmann claimed that “he did his duty [...] he not only obeyed orders, he also obeyed the law" (Arendt, 1963: 135) and as a consequence he was innocent. Arendt (1963) argued that the biggest crimes in history were implemented by ordinary people who were obedient to orders. Within social psychology, the paradigmatic case is provided by Milgram’s (1963, 1974) series of psychology experiments with deception and a strong authority presence verbally and with increasing pressure requiring subjects, for the sake of science, to press a button supposedly to implement an escalating series of electricity shocks to a confederate every time the latter provided a wrong answer: 62.5% of the subjects continued up to the maximum 450 volts electric shock.[28]

In most economic organizations, typically economic stakes rather than physical harm are involved and there is a much subtler role of authority. The desire for conformity may be seen to stem from maximizing the social image that the subject has with respect to the authority and, when conformity is to an anti-social act as harming others, may correspondingly minimize the moral costs from destruction, as the responsibility of being nasty is perceived to fall with the authority rather than oneself.[29] The outcome is the willingness to act aggressively

[27] For instance, Alwin (1990) showed that parents value obedience as one of the top three characteristics for a child to have.

[28] Similar results have been found under a range of variants and of subject pools (e.g., Kilham and Mann, 1974; Shanab and Yahya, 1977, 1978; Meeus and Raaijmakers, 1995; Blass, 1999).

[29] See Abbink and Herrmann (2011) for a discussion of the ‘moral costs of being nasty’.

towards other individuals because of obedience to the cue provided by the authority.[30]

There is a natural way to implement authority in the laboratory, and that is by exploiting the authority of the experimenter: subjects view the experimenter as being in a position of authority due to its legitimacy and expertise about the experimental environment (Orne 1962, 1973; Rosnow and Rosenthal 1997; Zizzo, 2010). Which is to say that (like Milgram, if more subtly) we use experimenter demand as our tool to study obedience: it is, in other words, for us precisely the object of investigation that we ride and look at the effects of rather than a confound to be isolated and controlled for. Indeed one motivation of our experiment, if of interest primarily only for experimental economists, is methodological: that is, to verify the extent to which experimental demand *can* affect behavior. Our main motivations however are more general: we look at *to what extent* authority is effective, even in a context where it implies going against standard social norms of norms and inducing damage to other people at no economic benefit (and indeed some economic cost) to one's own; and, relatedly, at whether we can shed at least some light on the factors that induce more or less obedience. Factors we consider are the effect of having more pressure to obey at given intervals of time, the effect of having an explicit reason to destroy and the expectation that the partner may damage back.

Our experiment employs the 'Joy of Destruction game' of Abbink and Herrmann (2011) as our baseline, where subjects are given the option to destroy half of the earnings of another subject at a price with no material benefit. Predictions can then be drawn straightforwardly using a simple model incorporating both social image and social preferences considerations. Our qualitative results are comparable to those of Abbink and Herrmann (2011), and, while arguably creating

[30] This would explain the fact that subordinates in organisations may not worry about the ethical implications of their actions if cued by the authority (Ashford and Anand, 2003; Darley, 2001), e.g. becoming willing to engage in race discrimination (Brief et al., 1995).

already some indirect demand to destroy, repeating the question 10 times does not make any statistical difference, with destruction rates still lower than 20%. However, when indirect pressure is reinforced by an explicit demand in the instructions to reduce their partners’ income, even if there is no explicit reason provided, the destruction rates more than doubles to over 40%. Asking subjects to destroy in specific intervals in time increases destruction rates further to around 60%, a result robust to providing an explicit justification for destruction or removing the potential for reciprocal aggression. Additional information is provided by qualitative data asking about the perceived objective of the experiment and by the use of a psychological social desirability scale that measures the extent of sensitivity to social pressure (Zizzo and Fleming, 2011; Stober, 2001). Overall, the evidence supports an explanation based on social image utility towards the authority (and corresponding moral costs from destruction being minimized).

There are a number of other connected strands of research. One is about experimenter demand characteristics by which a number of experiments are criticized on the grounds that experimental subjects may change their behavior due to implicit cues about what constitutes appropriate behavior (Orne, 1962; Rosnow and Rosenthal, 1997; Zizzo, 2010). Other related literature is where information is provided on others’ behavior e.g. in the context of giving (Cason and Mui, 1998; Frey and Meier, 2004; Landry et al., 2006; Krupka and Weber, 2008; Shang and Croson, 2009; Bicchieri and Xiao, 2009) or of public good contribution (Bardsley and Sausgruber, 2005), or where there are exogenous recommendations given in contexts where it can serve one’s own self interest, such as threshold public good games (Croson and Marks, 2001) or as a device to solve coordination in Chicken games (Cason and Sharma, 2007; Duffy and Feltovich, 2008). Cadsby et al. (2006) comes perhaps closest to our experiment in requiring subjects to pay a cost in a pro-social context; their manipulation is stronger than ours in that subjects are not simply requested but also *expected* to act in line with the request of the experimenter authority; subjects are given an explicit reason to give which is to help fund future experiments. Counter to the usual emphasis of behavioral economists on cooperation, there is a small but growing body of research on antisocial behavior to which this chapter is also related to.[31] This research finds that there are conditions under which subjects are willing to pay money to damage other's earnings, even at a cost to their own (Zizzo and Oswald, 2001; Zizzo, 2003; Abbink and Sadrieh, 2009; Abbink and Herrmann, 2011; Abbink et al., 2010).[32] There are also conditions in which 'antisocial' punishment takes place (Nikiforakis, 2008, Nikiforakis and Engelmann, 2008, Herrmann et al., 2008 and Denant-Boemont et al., 2007).

Section 2 describes the game and considers how different considerations such as social image towards the authority and peers, anticipated reciprocity, inequality aversion and spite affect its predictions. Section 3 provides the experimental design and specific hypotheses in the light of it and of Section 2. Section 4 presents the results, section 5 some supplementary analysis and section 6 the discussion and conclusions.

2. The Joy of Destruction Game with Obedience

The One Sided Joy of Destruction Game

There are two players, X and Y, with payoffs respectively x and y. X is sent the cue by the authority to destroy. If X chooses to destroy, X loses 1 Guilder and Y loses 5 Guilders, which means an advantageous inequality (i.e. $x > y$) equal to 4, and the indicator variable I_{DX} is set equal to 1; if not, it is equal to 0. If there is disadvantageous inequality, i.e. $x < y$, then the indicator variable I_i is equal to 1; if not, it is equal to 0; this case cannot occur in this one sided game. If there is

[31] Herrmann and Orzen (2008) refer to this as *homo rivalis* behavior.

[32] Abbink and Herrmann's (2011) idea that subjects are more willing to engage in antisocial behavior if the moral cost of it is lower has some parallelism with the 'moral wriggle room' literature (e.g., Dana et al., 2007).

advantageous inequality, which will always occur in this version of the game if the player obeys, then the indicator variable I_j is equal to 1; if not, it is equal to 0. Define, as per Fehr and Schmidt (1999), the weight on disadvantageous inequality s_i as larger (or at least as large) than that on advantageous inequality s_j: that is, $s_i \geq s_j$. We also impose the restrictions that $s_i > 0$ (disadvantageous inequality hurts) and $s_i \geq |s_j|$. This second restriction allows for spite in the sense that it is possible for the weight on advantageous inequality to be negative, and simply states that players would not be more spiteful when being ahead than when being behind. When positive, s_i and s_j can also be seen to potentially reflect, in reduced form, any reciprocity concern that X may have towards Y, without needing additional terms in the utility function; we can see this already in this one sided game for the case of s_j, since the only case that advantageous inequality can occur in this game is if X has been unkind to Y and may feel guilty as a result.[33]

We want to model social image concerns, which could equivalently be seen to reflect the moral costs of nastiness towards either the authority or the other player (the partner).[34] If X obeys, X gets social image warm glow with respect to the authority, i.e. obedience utility s_D; if not, X gets social image warm glow with respect to the partner s_N. s_D could also be seen to model any pure joy of destruction motive (Abbink and Herrmann, 2011) that the player has. We restrict s_D and s_N to be non-negative, i.e. s_D, $s_N \geq 0$. (s_D - s_N) can be interpreted as the net social image utility of destruction. We can now write down X's utility function U_X as follows:

[33] For a sophisticated and general guilt aversion model, see Battigalli and Dufwemberg (2009). Charness and Rabin (2002), Falk and Fischbacher (2006) and Cox et al. (2007) are examples of social preferences models allowing for reciprocity.

[34] To clarify this, each action can be seen as associated to a moral cost. This moral cost is such that X gets utility s_D if X obeys, in the sense that s_D is the social image utility net of the moral cost of obedience. Similarly, the moral cost from not obeying is such that X gets utility s_N by not obeying. Note also that we do not exclude the possibility that s_D may reflect a rule of thumb of relying on the cue provided by the authority.

(1) $U_X = x + I_{DX} s_D + (1 - I_{DX}) s_N - I_i s_i (y - x) - I_j s_j (x - y)$

Claim 1. X obeys and destroys if and only if the net social image utility of destruction and/or if spite is sufficiently high.

Proof: Define a threshold variable $\tau_1 = s_D - s_N - 4s_j - 1.$ Given (1) and the game parameters, obedience is chosen iff $\tau_1 \geq 0,$ which, in order to hold, requires either a sufficiently high (s_D - s_N) or a sufficiently negative s_j weight or a combination of the two.

The Two Sided Joy of Destruction Game

This is the game used by Abbink and Herrmann (2011). Both players simultaneously decide whether or not to destroy 5 Guilders of the partner's endowment, at an own cost of 1 Guilder. Define I_{DY} an indicator variable equal to 1 if player Y chooses to destroy (else 0). The utility function of player Y can be written, symmetrically to (1), as:

(2) $U_Y = y + I_{DY} s_D + (1 - I_D) s_N - I_i s_i (x - y) - I_j s_j (y - x)$

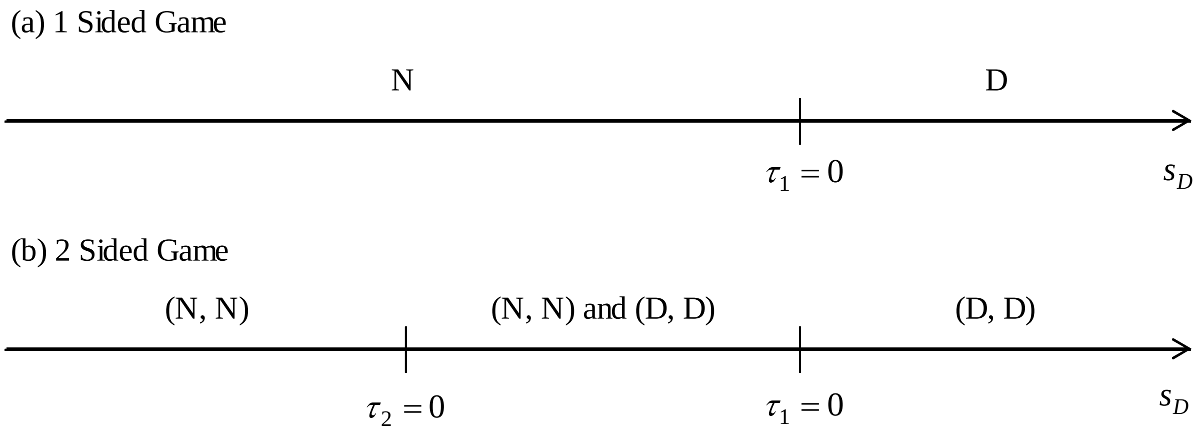
Given the experimental parameters and utility functions (1) and (2), Table 1 represents the game payoff matrix, labeling the actions as either D (Destroy) or N (Not Destroy), and Figure 1 represents the Nash equilibria of the game in the light of Claim 2, as stated next.[35] (N, D), the only case where player X ends up with less than Y, is also the only case where X is kind to Y while Y is being unkind back[36] and so the s_i weight can be postulated to incorporate also negative reciprocity

[35] In what follows we focus on pure strategy Nash equilibria only; mixed strategies equilibria are implausible here, bearing in mind that, even in treatments where (as described later) there are repeated requests to destroy, a decision to destroy can only be taken once and concludes the game.

[36] The symmetrical claim applies to Y in relation to outcome (D, N).

concerns as well as the disutility from disadvantageous inequality. Define the threshold variable $\tau_2 = s_D - s_N + 4s_i - 1$.

Figure 1. Predictions in the Joy of Destruction Game



(a) 1 Sided Game

(b) 2 Sided Game

Notes. N stands for Not destroy and D stands for Destroy. s_D is the social image utility towards the authority from destruction (inclusive of any joy of destruction). τ_1 and τ_2 are threshold variables as defined in the main text**.**

Claim 2. If $\tau_1, \tau_2 > 0,$ (D, D) is the strictly dominant strategy. If $\tau_1 \le 0$ and $\tau_2 \ge 0,$ both (D, D) and (N, N) are Nash equilibria. If $\tau_1, \tau_2 < 0,$ (N, N) is the strictly dominant strategy. Therefore, destruction is chosen if and only if the net social image utility from obedience and/or if spite is sufficiently high, whereas non destruction is chosen if and only if the net social image utility from obedience is sufficiently low and/or the weight on disadvantageous inequality is sufficiently high.

Proof: Trivial given the payoff matrix in Table 1, from which immediately follows that (N, N) is a Nash equilibrium iff $\tau_1 \le 0,$ while (D, D) is a Nash equilibrium if $\tau_2 \ge 0$ and a set of equilibria where both conditions hold is non-empty since $\tau_1 \ge \tau_2$ always as $s_i \ge |s_j|$. Also note that iff $\tau_1, \tau_2 < 0$ players strictly prefer N regardless the partner's choice, thus ensuring that (N, N) is strictly dominant; equally, $\tau_1, \tau_2 > 0$ means that players strictly prefer D regardless of the partner, thus ensuring strict dominance of (D, D). The second part of the claim just puts the equilibrium conditions into words.

Claim 3. Under pure self-interest, players do not destroy in both the one sided and two sided Joy of Destruction Game.

Proof. This follows from Claim 1 and Claim 2 by setting $s_D = s_N = s_i = s_j = 0.$

We now want to try to formalize the intuition that, in the two sided game, we may expect higher destruction than in the one sided game again because of anticipated reciprocity, i.e. because each player expects the partner to destroy his or her earnings.

Claim 4: If players expect partners to destroy, destruction is an equilibrium strategy in the two sided Joy of Destruction game even for parameter ranges where destruction is not a chosen strategy in the one sided Joy of Destruction game.

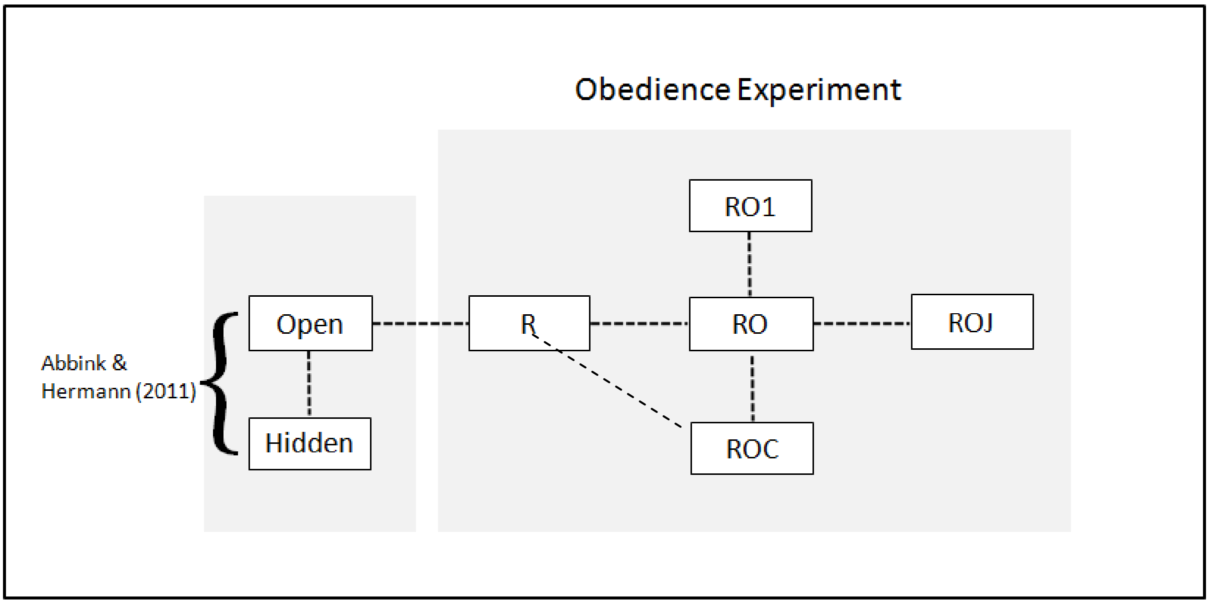
Proof: Claim 1 implies that obedience is never observed for $\tau_1 < 0$ whereas Claim 2 implies that also in the range $(\tau_1 < 0,\ \tau_2 \geq 0)$ (D, D) is an equilibrium.

We note, however, that in the range $(\tau_1 < 0,\ \tau_2 \geq 0),$ (N, N) is *also* an equilibrium. This is a useful point because it shows that, even for players caring about inequality aversion or reciprocity, it is not necessarily the case that behavior will be different in the two games.

The Repeated Game

As described below, the repeated game (either one sided or two sided Joy of Destruction game) simply gives players who have not destroyed before an opportunity to destroy again. Subjects are explicitly told that the number of rounds is fixed and that they receive no feedback on what the partner has done, and vice versa. As such, no punishment strategy can be exploited and simple backwards induction shows that the claims above still apply.

Figure 2. Experimental Design



Notes. Each box represents an experimental treatment. R stands for Repeated, RO for Repeated Obedience, ROC for Repeated Obedience Constant pressure, RO1 for Repeated Obedience 1 sided and ROJ for Repeated Obedience Justified.

3. Experimental Design and Hypotheses

Outline

The experiment was conducted at the University of East Anglia between January and March 2012 with 350 subjects. The participants were mostly students with a variety of different backgrounds. The experiment was in paper and pencil. The instructions were as close as possible to those of Abbink and Herrmann (2011). The experiment employed a fictional currency, called *Guilders*, which was converted to pounds at the end of the experiment at the rate of £0.75 per Guilder. Each session lasted approximately 60 minutes and the subjects earned on average £8.06 (approximately 12.79 US dollars), including a show-up fee of £2.00. Earnings were paid privately and anonymously at the end of the experiment. Subjects were not allowed to participate in more than one session. Each session lasted approximately one hour. The experiment consisted of six treatments (see Table 1), described below: Open (O), Hidden (H), Repeated Interaction (R), Repeated Obedience (RO), Repeated Obedience Justified (ROJ), and Repeated Obedience 1-sided (RO1). We ran 18 sessions in total.[37]

The Open and Hidden Treatments (O and H)

We began with a straightforward replication of Abbink and Herrmann's (2011) two treatments. The O treatment was exactly as the two sided Joy of Destruction game described in section 2, except that there was no cue by the authority/experimenter. Two players were endowed with 10 Guilders each, and both players simultaneously decided whether or not to destroy 5 Guilders of the other player's endowment, at an own cost of 1 Guilder. If they both reduced their

[37] We aimed for (at least) 40 independent observations per treatment, which meant 40 subjects for the O, H and R treatment, and 80 subjects for the RO1 treatment (since, as discussed later, only half of the subjects made actual destruction decisions); as we were able to have a few more subjects, we had 56 subjects for the RO treatment and 54 for the ROC treatment.

partners' income, they both earned 4 Guilders. If one reduced his/her partner's income but the other did not, the first earned 9 Guilders and the second earned 5 Guilders (and vice versa). If no one reduced the other person's income they both earned 10 Guilders.

In the H treatment, a die was also rolled for each player. If it turned out 1 or 6 (1/3 probability) the player would lose 5 Guilders regardless of the other player's decision; if the die turned out 2, 3, 4, or 5 (2/3 probability) the partner's decision was implemented. A player who lost 5 Guilders was not told whether this was due to the partner's action, or because of the roll of the die. Abbink and Herrmann (2011) predicted that in the H treatment the destruction rate would be higher than in the O treatment, as the moral costs of nastiness would be decreased by the player being able to hide behind the possibility of destruction by nature and being able to reason that, had the player not destroyed, destruction may have occurred anyway. Define *destruction rate* the proportion of subjects who choose to destroy. Formally, in the H there is less of a social image utility gain from choosing N (not destroying) relative to the O treatment, i.e. s_N is lower, increasing the parameters ranges where (D, D) is an equilibrium and (N, N) is not an equilibrium as per Claim 2. In Figure 2, this corresponds to a move of the $\tau_1 = 0$ and $\tau_2 = 0$ thresholds to the left.

As in Abbink and Herrmann (2011), after subjects had decided if they want to reduce their partner's income or not, we used an incentivized questionnaire in which we asked the participants about their expectation of their partner's behavior (i.e. their choice whether to destroy or not). If their prediction was correct they were rewarded with 1 Guilder.

After the incentivized questionnaire and before subjects were informed about their earnings we requested them to complete two questionnaires. The first was a Social Desirability Scale questionnaire (Stöber, 2001) and the second collected demographic information.

The Repeated Interaction Treatment (R)

The remaining treatments were identical to the O treatment except where otherwise specified below. In the Repeated Interaction (R) treatment the subjects were informed that the experiment consists of a predetermined amount of rounds although they will not be informed about the number of rounds until the end of the experiment. There were ten rounds.

At the start of the first round subjects were asked if they wanted to reduce their partner's income at the cost of 1 Guilder. After all subjects made their choices, the experimenters would record their choices in a separate sheet of paper. If they decided to reduce their partner's income, they did not make any further decision within the game, i.e. in the remainder of the 10 rounds. If they decided *not* to reduce their partner's income, in the following round they were asked if they were sure that they did not want to change their choice (i.e. reduce their partner's income). This question was posed to the subjects until they either changed their choice or until the experiment reached the final round. No feedback on their choices was provided to, or received about the actions of their partners, in between rounds.

Note that subjects could only destroy once throughout the game; once they chose to destroy, the game was effectively completed as far as they were concerned and they would have to wait.[38]

The aim of this treatment was to control for the effect of having repeated rounds. If this has any effect, it can be expected to be in the direction of increasing destruction. This is because it seems to starts implicitly to build up on the pressure to destroy by repeating the question ten times. If effective, this mechanism would

[38] We asked subjects whether they would have been willing to reduce their partners' income for an additional 5 Guilders at the cost of 1 Guilder, but this was a purely hypothetical and unincentivized question. See appendix for a brief analysis of the hypothetical responses

operate by raising s_D in our model, increasing the parameters ranges where (D, D) is an equilibrium and (N, N) is not an equilibrium.[39]

Hypothesis 1: A higher destruction rate is expected in the R treatment than in the O treatment.

Repeated Obedience (RO)

The Repeated Obedience (RO) treatment extends the R treatment. In this treatment, rounds 1, 4, 7 and 10 were marked with decision sheets provided on yellow paper. In this treatment the participant were told: “in the rounds with yellow instructions it would be especially useful if you were to reduce your partner’s income if you have not done so already. You are entirely free not to reduce if you wish.”

This was the extent of the experimenter cue for the subject to obey. It was deliberately a subtle cue for three reasons. First, in many real world settings commands are phrased in similar equivalent subtle language, and we wanted our experiment to be applicable to more than contexts where a direct command is given. Second, and relatedly, by showing the impact of having a subtle cue, we are identifying lower bounds to what may be the real world effectiveness of direct commands. Third, when authority is implemented as an experimental demand as we do, a direct command may be problematic, since subjects may then be confused by the instructions in believing that they have no option but to destroy, thus confounding the results. We made instead explicit that subjects were entirely free not to destroy if so they wished, so as to dispel any potential confusion on the matter.

[39] While the decision problem is simple, additional destruction may also be due to any decision error that may occur, and this bias is also picked up by this treatment. Subjects who by mistake do not destroy can (up to round 9) correct their mistake in the following round or rounds. A decision to destroy instead cannot be reverted.

Unlike Milgram (1963, 1974) or Cadsby et al. (2006), we did not provide an explicit reason to destroy in this treatment, though our instructions focused subjects on destroying as being particularly useful in the ‘yellow rounds’. As part of the end of experiment questionnaires, we did however add an open ended question, in which we asked the subjects what they believed was the scientific objective of the experiment. This question was provided in all the following treatments as well.

We expect subjects to destroy more in the experiment as a result of the experimental authority cue being provided. Obedience to the cue, by destroying (particularly in the ‘yellow rounds’), brings a higher utility s_D out of social image towards the authority (or, equivalently, lower moral cost of destruction). Alternatively or in combination, if subjects anticipate obedience on the part of their partners, this could lead them to a (D, D) rather than a (N, N) equilibrium in the middle range of Figure 1 where both equilibria exist.

Hypothesis 2: A higher destruction rate is expected in the RO treatment than in the R treatment.

The remaining treatments extend the RO treatment in various ways and help us shed light on why, if Hypothesis 2 receives support, subjects appear willing to destroy half of their partner’s earnings. Note, however, that there is no reason to expect inequality aversion, pure spite or joy of destruction to be different in this treatment from the previous ones, and so any change in destruction rate can be ascribed to the cue motivating subjects to obey via the increase in s_D, or alternatively (or in combination) leading to anticipated reciprocity.

Repeated Obedience Constant Pressure (ROC)

The Repeated Obedience Constant pressure treatment (ROC) differs from the RO treatment in that the participants were told that "it would be especially useful if you were to reduce your partner's income if you have not done so already, you are entirely free not to reduce if you wish". Rather than providing peak pressure to destroy at given points in time, a constant pressure to obey is provided over the ten rounds. We are therefore able to control for the effect of having 'yellow rounds'.

We expect the destruction rate to be intermediate between the R treatment and the RO treatment. This is because there is a cue by an authority to destroy but without the peak psychological pressure, and corresponding social image utility s_D from destroying (and/or anticipated reciprocity), that is provided by having certain rounds as 'yellow rounds'.

Hypothesis 3: A higher destruction rate is expected in the ROC treatment than in the R treatment.

Hypothesis 4: A higher destruction rate is expected in the RO treatment than in the ROC treatment.

Repeated Obedience Justified (ROJ)

The Repeated Obedience Justified treatment (ROJ) differs from the RO treatment in that, in addition to the same cue as in the RO treatment, participants are told that reducing their partner's income "would help us achieve a scientific objective of the experiment". Providing an explicit reason for destruction should increase destruction by increasing the moral legitimacy or providing an excuse, either way therefore increasing further the social image utility s_D from obeying to the authority (and/or increasing the likelihood of anticipated reciprocity). We should, of course, also still expect destruction to be higher than in the R treatment.

Hypothesis 5: A higher destruction rate is expected in the ROJ treatment than in the RO treatment.

Hypothesis 6: A higher destruction rate is expected in the ROJ treatment than in the R treatment.

Repeated Obedience 1-sided (RO1)

The Repeated Obedience 1-sided treatment (RO1) differs from the RO treatment in that half of the subjects were active and half were passive (though this terminology was not used in the instructions). Each active subject was matched with a passive partner but only the active subject made destruction decisions within the game, and this was known. Specifically, active subjects were told that "your partner answers some hypothetical questions but makes no decisions affecting your or his or her earnings".

The RO1 treatment tests whether the cue by the authority induces destruction because of anticipated reciprocity. If destruction is entirely driven by an unconditional desire to obey out of social image concerns as modeled in a higher s_D, there is no reason to expect RO1 destruction to differ from RO destruction. This is because the threshold τ_1 above which destruction is strictly dominant and so independent of expectations about the partner in the two sided game, as per Claim 2, is the same threshold above which obedience takes place in the one sided game as per Claim 1.

Hypothesis 7: The same destruction rate is to be expected in the RO and RO1 treatments.

We have shown with Claim 4, however, how anticipated reciprocity potentially triggers destruction under a wider range of parameter combinations in the two sided game than in the one sided game. This is because, within the range

$(\tau_1 \leq 0,\ \tau_2 \geq 0)$, (D, D) is an equilibrium enforced by the belief that the partner will destroy.

Hypothesis 8: A higher destruction rate is expected in the RO treatment than in the RO1 treatment.

As no anticipatory retaliation was possible in this treatment, there was no question regarding expected destruction from the partner at the end of the experiment.

4. Results

Figure 3. Destruction Rates accross Treatments

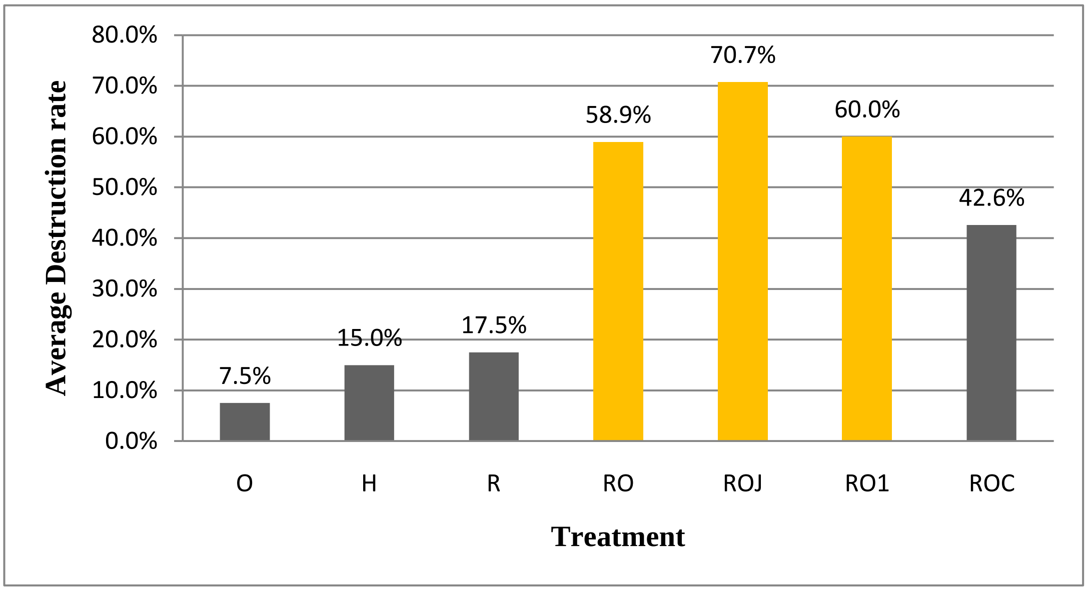


Figure 3 presents the average destruction rate across treatments. Table 2 presents the results of Probit regressions on whether subjects choose to destroy (=1) or not (=0) and their implications in terms of overall winning probabilities for each treatment. The regressions employ dummy variables for the experimental treatments; the RO treatment was used as baseline. SocialDesirability is our questionnaire social desirability measure for sensitivity to social pressure, and some demographic variables are included as controls. We now consider the hypotheses and the key evidence.

Table 2. Regressions on destruction rate

	Regression 1			Regression 2		
	b	se	p	b	se	p
O	-1.789****	0.37	0.000	-1.732****	0.37	0.000
H	-1.413****	0.30	0.000	-1.430****	0.31	0.000
R	-1.220****	0.30	0.000	-1.120****	0.30	0.000
ROJ	0.219	0.27	0.422	0.249	0.27	0.365
RO1	-0.058	0.27	0.831	-0.046	0.27	0.868
ROC	-0.493**	0.25	0.049	-0.475*	0.25	0.060
SDS17 Score				0.053**	0.02	0.041
British	-0.290	0.21	0.169	-0.249	0.21	0.245
Chinese	-0.216	0.26	0.406	-0.293	0.27	0.282
Gender	0.227	0.16	0.162	0.194	0.16	0.653
Age	-0.008	0.02	0.661	-0.007	0.02	0.688
Economics	-0.186	0.22	0.403	-0.148	0.03	0.505
Christian	-0.429*	0.31	0.098	-0.368	0.04	0.236
Atheist	-0.496	0.30	0.136	-0.440	0.05	0.136
Muslim	0.006	0.46	0.990	0.002	0.06	0.997
Constant	0.813**	0.35	0.021	0.746**	0.35	0.032
N	310			310		
Pseudo R-sqr	0.199			0.207		
Prob > X^2	0.000			0.000		

	Marginal Effects given Regression 1			Marginal Effects given Regression 2		
	b	se	p	b	se	p
O	0.074*	0.04	0.076	0.081*	0.04	0.070
H	0.139***	0.05	0.008	0.134***	0.05	0.008
R	0.185***	0.06	0.003	0.187***	0.06	0.003
RO	0.611****	0.06	0.000	0.601****	0.06	0.000
ROJ	0.690****	0.07	0.000	0.690****	0.07	0.000
RO1	0.589****	0.08	0.000	0.584****	0.07	0.000
ROC	0.422****	0.07	0.000	0.422****	0.06	0.000

Notes: Probit regressions with robust standard errors. * p < 0.1, ** p<0.05,*** p<0.01, **** p<0.001. Other than treatment variables, regressions include age (subtracted from mean age), gender (=1 for women), economics background (=1 if applicable), nationality (British=1 for British subjects, and Chinese =1 for Chinese subjects), religion (Christian = 1 for Christian subjects, Atheist =1 for atheist and agnostic subjects and Muslim = 1 for Muslim subjects). SocialDesirability includes a measure of social desirability.

Result 1: Against Hypothesis 1, there is no statistically significant difference between destruction rates in the O and the R treatments.

Evidence: R destruction is roughly the same as those in the H treatment and both are qualitatively above O, which is what we would expect. However, a Wald test for whether the coefficient on R is the same as the coefficient on O is not statistically significant for both Regression 1 ($p = 0.15$) and 2 ($p = 0.18$). The same result can be obtained in a simple bivariate test by comparing destruction rate proportions in the two treatments using a Fisher's exact test ($p = 0.16$).[40] Overall, repeating the question again and again does not make much difference, with the destruction rate remaining below 20%.

Result 2: In support of Hypothesis 2, the destruction rate is statistically significantly higher in the RO treatment than in the R treatment. The difference is large.

Evidence: Figure 2 shows how the destruction rate more than triples in moving from the R treatment (17.5%) to the RO treatment (58.9%). This is significant in a bivariate Fisher's exact test ($p < 0.001$). In the regression analysis, the coefficients on R are negative and statistically significant in both regressions 1 ($p < 0.001$) and 2 ($p < 0.001$).

Result 3: In support of Hypothesis 3, the destruction rate is statistically significantly higher in the ROC treatment than in the R treatment. The difference is large.

Evidence: Figure 2 shows how the ROC destruction rate is intermediate (42.6%) between R and RO, but more than double than that in R. A bivariate

[40] All p values reported in this paper are two tailed.

Fisher’s exact test is significant (p = 0.04) and, in our regression analysis, Wald tests also achieve significance in both regressions 1 (p = 0.02) and 2 (p = 0.01).

Result 4: In support of Hypothesis 4, the destruction rate is statistically significantly higher in the ROC treatment than in the RO treatment.

Evidence: As shown by Figure 3, quantitatively, the destruction rate in RO is about 16% more than in ROC, suggesting that providing peak pressure at intervals makes a difference. A bivariate Fisher’s exact test yields p = 0.06. In the regression analysis, controlling for SocialDesirability the coefficient on the ROC dummy has p = 0.06 (Regression 2); without this variable, which may be picking up some of the extra effect in RO, we have p = 0.049 (Regression 1). Although obviously less overwhelming than for Result 2 and 3, and bearing in mind that these are two tailed p tests whereas Hypothesis 4 is one-tailed, there is support for peak pressure at intervals having induced greater destruction. The supplementary analysis of section 4.2 will provide complementary evidence.

Result 5: Against Hypothesis 5, there is no statistically significant difference between destruction rates in the RO and the ROJ treatments.

Evidence: While the destruction rate is as high as 70.7% in the ROJ treatment, the difference from the RO treatment is not enough to achieve statistical significance either in a bivariate Fisher’s exact test (p = 0.18) or in the regression analysis by looking at the coefficient on the ROJ dummy (p = 0.42 in Regression 1 and p = 0.37 in Regression 2). Providing an explicit reason for destruction does not seem to make a difference.

Result 6: In support of Hypothesis 6, the destruction rate is statistically significantly higher in the ROJ treatment than in the R treatment. The difference is large.

Evidence: The destruction rate is about five times as high in the ROJ treatment as in the R treatment. In the light of Result 3, it should then come to no surprise that this difference is statistically significant in a bivariate Fisher's exact test ($p < 0.001$) or in the regression analysis, where Wald tests for whether the coefficients on the ROJ dummy are equal to those on the R dummy yield $p < 0.001$ for both Regressions 1 and 2.

Result 7: In support of Hypothesis 7 and against Hypothesis 8, there is no statistically significant difference between destruction rates in the RO and the RO1 treatments.

Evidence: Figure 3 already provides the answer by showing virtually identical destruction rates (60% in RO1 vs. 58.9% in RO), not statistically significantly different in a bivariate Fisher's exact test ($p = 0.51$). The insignificance of the RO1 dummy in Regressions 1 ($p = 0.83$) and 2 ($p = 0.87$) further confirm this. Overall, this points to obedience to authority rather than anticipated reciprocity by partners as the driver of destruction.[41]

[41] It is interesting to compare and contrast our simple between treatments test on anticipation of reciprocity with the use of belief elicitation data, which we also collected at the end of the experiment for all treatments other than RO1 (where it obviously could not be formulated). There tends to be a positive correlation between destruction and stated belief about the destruction of the other partner, though there is no clear pattern to it across treatments (Spearman $\rho = 0.854$, - 0.009, 0.223, 0.297, 0.480, 0.180, respectively with $p < 0.001$, $p =$ 0.96, 0.17, 0.03, < 0.01, and = 0.195 in treatments 0, H, R, RO, ROJ and ROC respectively). It is possible that having decided to destroy might make it more likely for subjects to believe the partner will, e.g. out of self-image and cognitive dissonance concerns.

Supplementary Analysis

The only other finding from Table 2 is that SocialDesirability predicts greater destruction (p = 0.04). The result supports the interpretation of destruction as obedience to social pressure from the authority. Demographic variables are generally insignificant.

Time of destruction. It is interesting to see when the decision to destroy takes place over the ten rounds of most treatments (i.e., all treatments other than O and H). Figure 4 presents destruction rates per round for the 'yellow rounds' repeated play treatments where peak pressure is applied to destroy at given points in time (RO, ROJ, and RO1) and for treatments where constant pressure was applied (R and ROC). Note that subjects can only destroy once, and therefore those who have already destroyed in the early rounds cannot destroy further in the following rounds; effectively, later decisions are conditional on not having destroyed before, reflecting a sample selection increasingly composed of subjects who are less willing to obey the experimental cue provided.

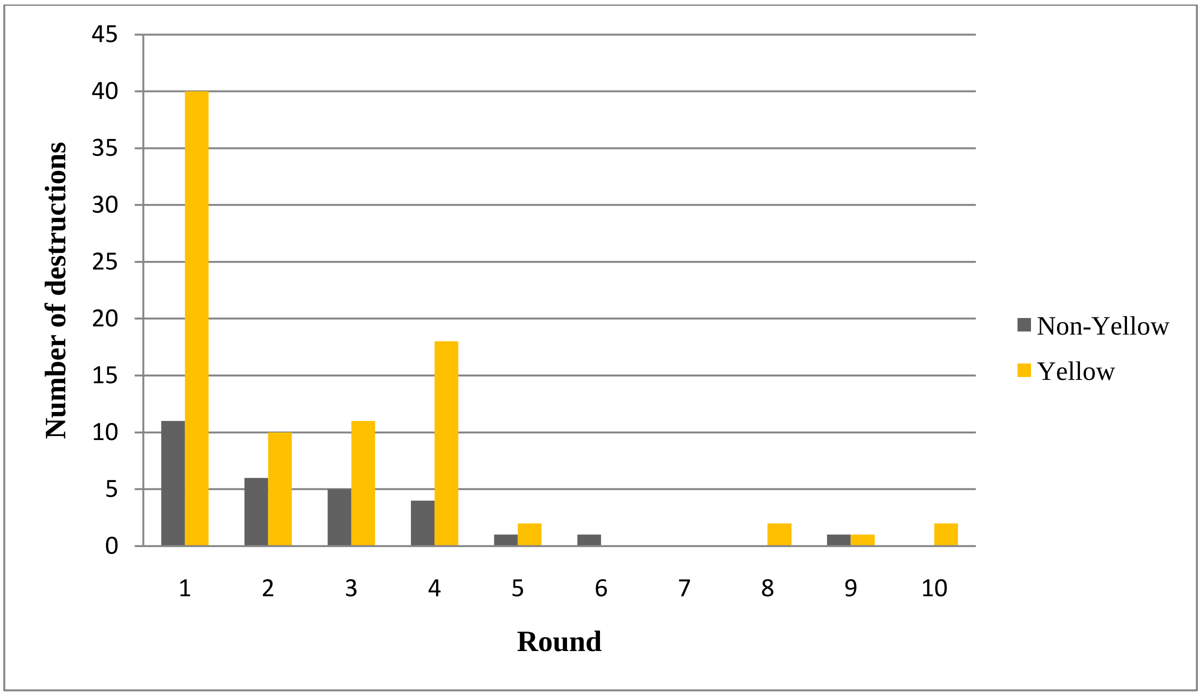
Table 3. Destruction rates in each round

Round	**R**	**RO**	**ROJ**	**RO1**	**ROC**
1	5.00%	25.00%	37.50%	25.00%	16.00%
2	2.63%	8.69%	15.38%	6.67%	11.11%
3	5.40%	9.52%	13.64%	14.30%	7.50%
4	0.00%	18.42%	26.31%	25.00%	10.81%

Notes: The destruction rate in round 1 is the proportion of all subjects who have destroyed in round 1. The destruction rates for rounds 2, 3 and 4 are conditional on the destruction in the previous rounds: specifically, they are the proportions of all subjects who have destroyed in the given round conditional on them not having destroyed in the previous round(s).

Figure 4 shows that almost all of the destruction takes places by the 4th round, and Table 3 provides a breakdown of destruction rates (conditional on the previous round's destruction from round 2 onwards) up to round 4. Despite the declining trend across rounds due to early decisions to destroy and sample selection, in the 'yellow rounds' treatments with peaks in pressure to destroy there is a spike in destruction rates in round 4. This is in line with our expectations, and complements Result 4 above, as round 4 was one of the rounds in which we asked the subjects explicitly to reduce their partners' income.

Figure 4. Destruction Rates per Round



Notes. Yellow stands for treatments with 'yellow rounds' RO, ROJ and RO1. Non-Yellow stands for treatments R and ROC.

Focusing on round 1, and in the lack of any peak pressure, round 1 destruction is not dissimilar between R and ROC and the one side and O (or H) on the other side. As subjects know that there are multiple rounds, subjects can, of course, choose to defer destruction to after round 1, and a majority does so, particularly though by no means exclusively in both R and ROC, the treatments with repeated but constant pressure to obey.[42]

Qualitative data. As noted earlier, in the treatments with an explicit cue by the authority/experimenter to destroy (RO, ROC, RO1 and ROJ), at the end of the experiment we asked subjects what they thought the objective of the experiment was.[43] The answers were then grouped in categories by research assistants themselves not informed of the objectives of the experiment.[44]

The answers of the subjects were divided into 7 broad categories: authority / effect of yellow instructions, willingness to reduce the others payoff at an own sacrifice, selfishness vs cooperation, trust, attitudes towards social and strategic interaction, and change in behavior over time, and rest.[45] Figure 5 displays the results of the classification.

Unsurprisingly, while between 15% and 30% of the subjects in the treatments with 'yellow rounds' peaks of pressure picked 'authority/effect of yellow instructions' as the objective of the experiment, less than 5% did so in the ROC

[42] We ran Probit regression analysis specifically on round 1 destruction (see appendix B). The coefficients on the O and R dummies are negative and statistically significant relative to the RO baseline; that on H is not.

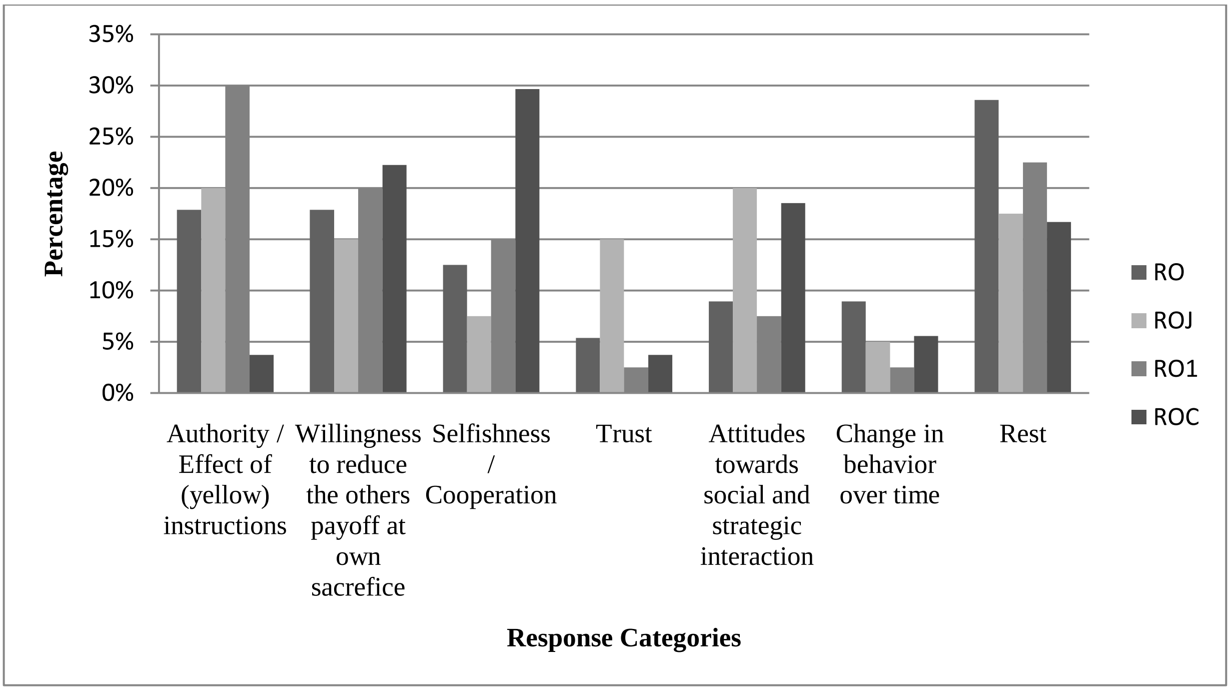
[43] Such data obviously has limitations, as discussed in Zizzo (2010).

[44] The answers were first grouped into categories by a research assistant with no prior knowledge of the aims of the experiment or experimental design. Afterwards a second research assistant, with no prior knowledge of the aims of the experiment, received the previously created seven categories along with the subjects' responses and was asked to independently match every answer with one of the categories. Finally, the assistants have met up to reconcile any discrepancy.

[45] *Authority / effect of yellow instructions* referred to subjects who believed that the aim of the experiment was either the influence of authority upon them or the influence of the 'yellow paper' rounds. *Willingness to reduce the others payoff at own sacrifice* saw the experiment being about subjects reducing their partner's income despite there was no monetary incentive for them to do so. *Selfishness / cooperation* referred to whether people are selfish or willing to cooperate. Similarly, *Trust* was about subjects being able to trust each other not to destroy. *Attitudes towards social and strategic interaction* was about when subjects argued that the aim of the experiment relates to strategic behavior, either on what the other subject will do or what he or she would do if he or she was in their place (RO1 treatment). *Change in behavior over time* answers were about the experiment being about change in choice if they are asked repetitively if they will change their choice. *Rest* was the residual category and included non-responders.

treatment with constant pressure and no 'yellow rounds'.[46] The modal answer in ROC switched to 'selfishness/ cooperation', again a plausible response insofar as subjects can avoid destroying each other.[47]

Figure 5. Proportion of qualitative responses per category for each treatment



Notes. The responses were to a question on what subjects thought was the objective of the experiment.

There seem to be three useful messages from the qualitative data. First, all of the six substantive categories pick up on potentially relevant aspects of the decision problem, displaying no evidence of confusion about the decision problem. Second, we had no evidence, either from these categories or looking at the residual category, of subjects wanting to destroy money in order to indirectly return money

[46] The difference between the ROC and the remaining treatments is statistically significant (Fisher exact tests, p = 0.017, 0.014 and < 0.001 for RO, ROJ and RO1 respectively).

[47] This difference is also statistically significant across all treatments (Fisher's exact tests, p = 0.024 p = 0.007 p = 0.078 for RO, ROJ and RO1 respectively).

to the experimenter, which would be a kind of house money effect. Third, there is no evidence that subjects wanted to be altruistic towards the experimenter as such; the focus even of the answers classified as related to ‘selfishness/cooperation’ was with respect to partners.

5. Discussion and Conclusions

Our results suggest that even a limited (rather than Milgram style) cue from the authority, without an explicit justification, can induce obedience on the part of some 60% of the subjects to halve the earnings of a partner at a cost to their own. A deliberate and explicit experimenter demand (Zizzo, 2010) was used as a tool to study the effect of authority, with the authority being the experimenter. In this sense, our study also checked how far one can go with experimenter demand when cues less explicit than making vocal demands to obey are used.

Our 60% destruction rate under obedience compares with previous experiments on antisocial behavior that have achieved destruction rates ranging from approximately 10-40% using methods such as the ability to hide due to random destruction (Abbink and Hermann, 2011; Abbink and Sadrieh, 2007) or the introduction of pointless prizes (Abbink and Hermann, 2009). Our results are made stronger by the fact that, while unequivocal, our cue was not phrased directly; indirect cues of this kind are arguably pervasive in workplaces and help us avoid the potential criticism that subjects did not understand that they had a choice not to destroy. Our results cannot be explained by purely repeating the task again and again, since by doing so destruction rates remain below 20%. They also cannot be explained by reciprocal expectations of destruction from the partner, since, even in the absence of the possibility of reciprocity, we observe around a 60% destruction rate. Providing pressure at peak intervals does help the authority to induce more obedience, but giving an explicit reason does not yield to a statistically significant increase. Our qualitative data suggests that subjects generally did understand the nature of the task and that they did not care about returning money to the experimenter. They do not more generally seem to suggest, within the range of motivations, that subjects wanted to be pro-social towards the experimenter as such. Of course, qualitative data has limitations and should be taken with caution.

Using our conceptual framework provided in section 2, changes in s_D, the social image towards the authority parameter, appear to drive the results. Because of the anti-social nature of the action to be undertaken by the agent, namely to destroy half of their partner's earnings, which goes against the standard social norm not to cause unnecessary harm to others, we can say that changes in s_D reflect social image concerns *towards the authority* as opposed to simply reflect an enhanced social norm. The cue by the authority minimizes the moral cost of destruction and enhances the desire to obey the authority as a way of maximizing social image utility. While this chapter has treated social image utility maximization and moral cost of destruction minimization as equivalent, and correspondingly modeled them with a single parameter, further research could try to disentangle the two. Overall, our experiment suggests that obedience should not be neglected in principal agent modeling, nor should it be neglected as a powerful managerial and social tool; in this sense, the emphasis by Cialdini and Goldstein (2004) on the role of compliance to authority in organizations appears well placed. It is arguable that, even in the lack of economic incentives, individuals may tend at least to some degree to obey orders, and that this is exploited by economic organizations big and small as a management tool. For example, it may help strategic delegation which is advantageous to principals in handling conflict and contests (Warneryd, 2012).

That said, where authority is in the wrong hands, cues to engage in aggressive or harassing behavior do not have to go as far as giving explicit orders. Of this too there seems to be evidence in organizations (Ashford and Anand, 2001; Brief et al., 2001; Darley, 2001). Saint Teresa of Avila (1972) may have been right in stating that obedience is powerful in making things easy which seem impossible; but the things it makes easy may or may not be socially desirable, and trying to understand how it operates appears relevant.[48] The converse of our finding, of course, is that some 30-40% of subjects were resistant to the cues by the authority, even when repeated ten times, under intervals of peak pressure and with an explicit justification, and this heterogeneity in the willingness to obey orders is something worth further attention too.

References

Abbink, K. and Hermann, B. 2011. "The Moral Costs of Nastiness." *Economic Inquiry*. 49, 631-634.

Abbink, K., and Sadrieh, A. 2009. *"*The Pleasure of Being Nasty.*"* *Economics Letters*, 105, 306-308.

Abbink, K., and Herrmann B. 2009. "Pointless Vendettas." Norwich: University of East Anglia CBESS Working paper 2009-02.

Abbink, K., and Herrmann B. 2011. "The Moral Costs of Nastiness." *Economic Inquiry*, 49, 631-633.

Abbink, K., Masclet, D., and van Veelen M. 2011. "Reference Point Effects in Antisocial Preferences." Montreal: CIRANO Discussion Paper 2011s-11.

Abbink, K., Brandts, J., Herrmann, B., and Orzen, H. 2010. "Inter-Group Conflict and Intra-Group Punishment in an Experimental Contest Game." *American Economic Review,* 100, 420-447.

[48] One specific further area for future research could revolve around understanding the interaction between social norms and obedience to authority. Another one could be the extent to which social image utility has been internalized by subjects as part of their self-image or is reliant on a simple rule of thumb. In both of these cases, social image utility would not depend on knowing that the authority may learn about their actions.

Alwin, D., 1990. "Cohort Replacement and Changes in Parental Socialization Values." *Journal of Marriage and Family*, 52, 347-360.

Andreoni, J., Bernheim, D. B. 2009. "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica*, 77, 1607-1636.

Arendt, H. 1963. *Eichmann in Jerusalem.* New York: The Viking Press.

Ariely, D., Bracha, A., and Meier, S. 2009. "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *American Economic Review*, 99, 544–555.

Asch, S. 1946. "Forming Impression of Personality." *Journal of Abnormal Psychology,* 41, 258–290.

Asch, S. E. 1955. "Opinions and social pressure." *Scientific American*, 193, 31-35.

Ashford, B. E, Anand, V. 2003. "The normalization of corruption in organizations." In Staw, B. M., and Kramer, R. M. eds., *Research in Organizational Behavior*, Greenwich, CT: JAI.

Bardsley, N., and Sausgruber, R. 2005. "Conformity and reciprocity in public good provision." *Journal of Economic Psychology*, 26, 664–681.

Battigalli, P., and Dufwemberg, M. 2009. "Dynamic Psychological Games." *Journal of Economic Theory*, 144, 1-35.

Benabou, R., and Tirole, J. 2006. "Incentives and prosocial behavior." *American Economic Review*, 96, 1652–1678.

Bicchieri, C., and Xiao, E. 2009. "Do the right thing: but only if others do so." *Journal of Behavioral Decision Making*, 22, 191–208.

Bicchieri, C. 2006. *The Grammar of Society: the Nature and Dynamics of Social Norms,* Cambridge: Cambridge University Press.

Blass, T. 1999. “The Milgram paradigm after 35 years: Some things we now know about obedience to authority.” *Journal of Applied Social Psychology, 29,* 955–978.

Brandts, J., Cooper, D., and Fatas, E. 2007. "Leadership and overcoming coordination failure with asymmetric costs." *Experimental Economics*, 10, 269-284.

Brief, A. P., Buttram, R. T., Elliot, J. D., Reizenstein, R. M., and McCline, R. L. 1995. “Releasing the beast: a study of compliance with orders to use race as a selection criterion.” *Journal of Social Issues*, 51, 177–194.

Cadsby, C., Maynes, E., and Trivedi, V. 2006. “Tax Compliance and Obedience to Authority at Home and in the Lab: A New Experimental Approach.” *Experimental Economics*, 9, 343-359.

Case, A. C., and Katz, L. F. 1991. “The company you keep: The effects of family and neighbourhood on disadvantaged youths.” NBER working paper 3705.

Cason, T., and Sharma, T. 2007. “Recommended Play and Correlated Equilibria: An Experimental Study.” *Economic Theory*, 33, 11–27.

Cason, T., and Mui, V. 1998. “Social influence in the sequential dictator game.” J*ournal of Mathematical Psycholog*y, 42, 248–265.

Cialdini, R. B., and Noah, J. G. 2004. “Social Influence: Compliance and Conformity.” *Annual Review of Psychology*, 55, 591-621.

Cox, J. C., Friedman, D., and Gjerstad, S. 2007. “A Tractable Model of Reciprocity and Fairness.” *Games and Economic Behavior*, 59, 17-45.

Croson, R., and Marks, M. 2001. “The effect of recommended contributions in the voluntary provision of public goods.” *Economic Inquiry, 39*, 238-249.

Dana, J., Weber, R., and Kuang, J. 2007. “Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness.” *Economic Theory*, 33, 67–80.

Darley, J. M. 2001. “The dynamics of authority in organization.” In Darley, J. M., Messick, D. M., and Tyler, T. R. eds. *Social Influences on Ethical Behavior in Organizations.* Mahwah, NJ/London: Erlbaum, 37–52.

Denant-Boemont, L., Masclet, D., and Noussair, C.N., N. N. 2007. “Punishment, Counterpunishment and Sanction Enforcement in a Social Dilemma Experiment.” *Economic Theory*, 33, 145-167.

Duffy, J., and Feltovitch, N. 2008. “Correlated equilibria, good or bad: An experimental study.” University of Pittsburgh and University of Aberdeen working paper.

Ellingsen, T., and Johannesson, M. 2008. “Pride and prejudice: The human side of incentive theory.” *American Economic Review*, 98, 990–1008.

Englmaier, F., and Wambach, A. 2010. “Optimal incentive contracts under inequity aversion.” *Games and Economic Behavior,* 69, 312-428.

Falk, A., and Fischbacher, U. 2006. “A Theory of Reciprocity.” *Games and Economic Behavior*, 54, 293-315.

Fehr, E., and Schmidt, K. 1999. “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics, 114,* 817-868.

Frey, B., and Meier, S. 2004. “Social Comparison and pro-social behavior: testing conditional cooperation in a field experiment.” *American Economic Review*, 94, 1717–1722.

Glazer, A., and Kai, A. K. 1996. “A Signaling Explanation for Charity.” *American Economic Review,* 86, 1019-1028.

Güth, W., Levati, M. V., Sutter, M., Van der Heijden, E. C. M. 2007. “Leading by example with and without exclusion power in voluntary contribution experiments.” *Journal of Public Economics*, 91, 1023-1042.

Herrmann, B., and Orzen, H. 2008. "The Appearance of Homo Rivalis: Social Preferences and the Nature of Rent Seeking." CeDEx Discussion Paper 2008-10.

Herrmann, B., Thöni, C., and Gächter, S. 2008. "Antisocial Punishment Across Societies." *Science,* 319, 1362-1367.

Jones, S. 1984. *The Economics of Conformism*. Basil Blackwell: Oxford.

Kawaguchi, D. 2004. "Peer effects on substance use among American teenagers." *Journal of Population Economics*, 17, 351-367.

Kilham, W., and Mann, L. 1974. "Level of Destructive Obedience as a Function of Transmitter and Executant Roles in the Milgram Obedience Paradigm." *Journal of Personality and Social Psychology*, *29,* 696–702.

Krupka, E., and Weber, R. 2009. "The focusing and informational effects of norms on pro-social behavior." *Journal of Economic Psychology*, 30 , 307–320.

Landry, C., Lange, A., List, J., Price, M., and Rupp, N., 2006. "Toward an understanding of the economics of charity: evidence from a field experiment." *Quarterly Journal of Economics*, 121, 747–782.

Lopez-Perez, Raul. 2008. "Aversion to Norm –Breaking: A Model." *Games and Economic Behavior*, 64, 237-267.

Lundborg, P. 2006. "Having the wrong friends? Peer effects in adolescent substance use." *Journal of Health Economics*, 25, 214-233.

Manski, C. 1993. "Identification of endogenous social effects: the reflection problem." *Review of Economic Studies*, 60, 531–542.

Meeus, W., and Raaijmakers, Q. 1995. "Obedience in Modern Society: The Utrecht Studies." *Journal of Social Issues*, *51*, 155–175.

Milgram, S. 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology*. *67*, 371–378.

Milgram, S. 1974. *Obedience to authority: an experimental view*. New York: Harper and Row.

Moxnes, E., Van der Heijden, E. 2003. “The effect of leadership in a public bad experiment.” *Journal of Conflict Resolution*, 47, 773–795.

Nikiforakis, N. 2008. “Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?” *Journal of Public Economics*, 92, 91-112.

Nikiforakis, N., and Engelmann, D. 2008. “Feuds in the Laboratory? A Social Dilemma Experiment", University of Melbourne Department of Economics Research Paper 1058.

Orne, M. T. 1962. “On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications.” *American Psychologist, 17*, 776-783.

Powell, L. M., Tauras, J. A., and Ross, H. 2005. “The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behaviour.” *Journal of Health Economics*, 24, 950-968.

Rosnow, R. L., and Rosenthal, R. 1997. *People Studying People: Artifacts and Ethics in Behavioral Research*. New York: Freeman.

Shanab, M., and Yahya, K. 1978. “A Cross-Cultural Study of Obedience.” *Bulletin of the Psychonomic Society,* 11, 267-269.

Shanab, M., and Yahya, K. 1977. “A Behavioral Study of Obedience in Children.” *Journal of Personality and Social Psychology,* 35, 530-536.

Shang, J., and Croson, R. 2009. “Field experiments in charitable contribution: The impact of social information on the voluntary provision of public goods.” *Economic Journal, 119*, 1422-1439.

Sliwka, D. 2007. “Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes.” *American Economic Review*, 97, 999–1012.

St. Teresa of Avila. 1972. *Interior Castle*. Image Books, New York.

Stöber, J. 2001. “The Social Desirability Scale-17 (Sds-17) - Convergent Validity, Discriminant Validity, and Relationship with Age.” *European Journal of Psychological Assessment*, 17, 222-232.

Van der Heijden, E., Potters, J., and Sefton, M. 2008. “Hierarchy and Opportunism in Teams.” *Journal of Economic Behavior and Organization*, 69, 39-50.

Zafar, B. 2011. “An Experimental Investigation of Why Individuals Conform.” *European Economic Review* 55, 774-798.

Zizzo, D. J. 2010. “Experimenter Demand Effects in Economic Experiments.” *Experimental Economics*, 13, 75-98.

Zizzo, D. J., and Fleming, P. 2011. “Can Experimental Measures of sensitivity to social pressure predict public good contribution?” *Economic Letters*, 111, 239-242.

Appendix

A. Experimental Instructions

Instructions (R Treatment)

Thank you for participating in this experiment. The experiment will last approximately 1 hour. If you read the following instructions carefully, you can, depending on your own decisions, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. It is not allowed to communicate with the other participants during the experiments. Should you have any questions, please ask us. If you violate this rule, we will have to dismiss you from the experiment and you will forfeit all payments.

You will be paid after the experiment. No other participant will know how much you earned. You will be paid £2 for showing up plus any additional earnings that you have in the experiment.

During the experiment you will have the chance to earn points, which will be converted into cash at the end of today's session, using an exchange rate of 1 Guilder = £0.75.

In the experiment you are randomly matched with another participant – your partner. You will not learn the identity of the participant you are matched with, and vice versa your partner will never learn about your identity.

The experiment has a fixed and predetermined number of rounds. You will not learn the outcome of the decisions of your partner until the end of the experiment, and vice versa your partner will not learn the outcome of your decisions until the end of the experiment.

You and your partner both receive an endowment of 10 Guilders. You then have to decide whether to reduce your partner's income or to leave it as it is. Reducing your partner's income will cost you 1 Guilder. By paying 1 Guilder, you can reduce the other partner's income by 5 Guilders. Your partner simultaneously takes the same decision. He can choose between leaving your income unaltered, or reducing it by 5 Guilders. Your partner will incur the same cost - 1 Guilders - if he or she chooses to reduce your income

If both of you choose to leave the other person's income unaltered, both of you will earn the 10 Guilders.

If both of you choose to reduce the other person's income, both of you will earn 4 Guilders (10-5-1).

If you choose to reduce your partner's income, but he/she decides to leave your income unaltered, you will earn 9 Guilders and your partner will earn 5 Guilders.

If you choose not to reduce your partner's income, but he/she decides to reduce yours, you will earn 5 Guilders and your partner will earn 9 Guilders.

Instructions (RO Treatment)

Thank you for participating in this experiment. The experiment will last approximately 1 hour. If you read the following instructions carefully, you can, depending on your own decisions, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. It is not allowed to communicate with the other participants during the experiments. Should you have any questions, please ask us. If you violate this rule, we will have to dismiss you from the experiment and you will forfeit all payments.

You will be paid after the experiment. No other participant will know how much you earned. You will be paid £2 for showing up plus any additional earnings that you have in the experiment.

During the experiment you will have the chance to earn points, which will be converted into cash at the end of today’s session, using an exchange rate of 1 Guilder = £0.75.

In the experiment you are randomly matched with another participant – your partner. You will not learn the identity of the participant you are matched with, and vice versa your partner will never learn about your identity.

The experiment has a fixed and predetermined number of rounds. You will not learn the outcome of the decisions of your partner until the end of the experiment. and vice versa your partner will not learn the outcome of your decisions until the end of the experiment.

You and your partner both receive an endowment of 10 Guilders. You then have to decide whether to reduce your partner's income or to leave it as it is. Reducing your partner's income will cost you 1 Guilder. By paying 1 Guilder, you can reduce the other partner's income by 5 Guilders. Your partner simultaneously takes the same decision. He can choose between leaving your income unaltered, or reducing it by 5 Guilders. Your partner will incur the same cost - 1 Guilders - if he or she chooses to reduce your income

If both of you choose to leave the other person's income unaltered, both of you will earn the 10 Guilders.

If both of you choose to reduce the other person's income, both of you will earn 4 Guilders (10-5-1).

If you choose to reduce your partner's income, but he/she decides to leave your income unaltered, you will earn 9 Guilders and your partner will earn 5 Guilders.

If you choose not to reduce your partner's income, but he/she decides to reduce yours, you will earn 5 Guilders and your partner will earn 9 Guilders.

Some of these predetermined rounds will be marked with yellow instructions; this is instructions in yellow paper. In the rounds with yellow instructions it would be especially useful if you were to reduce your partner’s income if you have not done so already. You are entirely free not to reduce if you wish.

Instructions (ROC Treatment)

Thank you for participating in this experiment. The experiment will last approximately 1 hour. If you read the following instructions carefully, you can, depending on your own decisions, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. It is not allowed to communicate with the other participants during the experiments. Should you have any questions, please ask us. If you violate this rule, we will have to dismiss you from the experiment and you will forfeit all payments.

You will be paid after the experiment. No other participant will know how much you earned. You will be paid £2 for showing up plus any additional earnings that you have in the experiment.

During the experiment you will have the chance to earn points, which will be converted into cash at the end of today’s session, using an exchange rate of 1 Guilder = £0.75.

In the experiment you are randomly matched with another participant – your partner. You will not learn the identity of the participant you are matched with, and vice versa your partner will never learn about your identity.

The experiment has a fixed and predetermined number of rounds. You will not learn the outcome of the decisions of your partner until the end of the experiment, and vice versa your partner will not learn the outcome of your decisions until the end of the experiment.

You and your partner both receive an endowment of 10 Guilders. You then have to decide whether to reduce your partner's income or to leave it as it is. Reducing your partner's income will cost you 1 Guilder. By paying 1 Guilder, you can reduce the other partner's income by 5 Guilders. Your partner simultaneously takes the same decision. He can choose between leaving your income unaltered, or reducing it by 5 Guilders. Your partner will incur the same cost - 1 Guilders - if he or she chooses to reduce your income

If both of you choose to leave the other person's income unaltered, both of you will earn the 10 Guilders.

If both of you choose to reduce the other person's income, both of you will earn 4 Guilders (10-5-1).

If you choose to reduce your partner's income, but he or she decides to leave your income unaltered, you will earn 9 Guilders and your partner will earn 5 Guilders.

If you choose not to reduce your partner's income, but he or she decides to reduce yours, you will earn 5 Guilders and your partner will earn 9 Guilders.

It would be especially useful if you were to reduce your partner’s income if you have not done so already. You are entirely free not to reduce if you wish.

Instructions (ROJ Treatment)

Thank you for participating in this experiment. The experiment will last approximately 1 hour. If you read the following instructions carefully, you can, depending on your own decisions, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. It is not allowed to communicate with the other participants during the experiments. Should you have any questions, please ask us. If you violate this rule, we will have to dismiss you from the experiment and you will forfeit all payments.

You will be paid after the experiment. No other participant will know how much you earned. You will be paid £2 for showing up plus any additional earnings that you have in the experiment.

During the experiment you will have the chance to earn points, which will be converted into cash at the end of today's session, using an exchange rate of 1 Guilder = £0.75.

In the experiment you are randomly matched with another participant – your partner. You will not learn the identity of the participant you are matched with, and vice versa your partner will never learn about your identity.

The experiment has a fixed and predetermined number of rounds. You will not learn the outcome of the decisions of your partner until the end of the experiment, and vice versa your partner will not learn the outcome of your decisions until the end of the experiment.

You and your partner both receive an endowment of 10 Guilders. You then have to decide whether to reduce your partner's income or to leave it as it is. Reducing your partner's income will cost you 1 Guilder. By paying 1 Guilder, you can reduce the other partner's income by 5 Guilders. Your partner simultaneously takes the same decision. He can choose between leaving your income unaltered, or reducing it by 5 Guilders. Your partner will incur the same cost - 1 Guilders - if he or she chooses to reduce your income

If both of you choose to leave the other person's income unaltered, both of you will earn the 10 Guilders.

If both of you choose to reduce the other person's income, both of you will earn 4 Guilders (10-5-1).

If you choose to reduce your partner's income, but he/she decides to leave your income unaltered, you will earn 9 Guilders and your partner will earn 5 Guilders.

If you choose not to reduce your partner's income, but he/she decides to reduce yours, you will earn 5 Guilders and your partner will earn 9 Guilders.

Some of these predetermined rounds will be marked with yellow instructions; this is instructions in yellow paper. In the rounds with yellow instructions it would be especially useful if you were to reduce your partner's income if you have not done so already. You are entirely free not to reduce if you wish. However if you do it would help achieving a scientific objective of the experiment.

Instructions (RO1 Treatment – active subjects)

Thank you for participating in this experiment. The experiment will last approximately 1 hour. If you read the following instructions carefully, you can, depending on your own decisions, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. It is not allowed to communicate with the other participants during the experiments. Should you have any questions, please ask us. If you violate this rule, we will have to dismiss you from the experiment and you will forfeit all payments.

You will be paid after the experiment. No other participant will know how much you earned. You will be paid £2 for showing up plus any additional earnings that you have in the experiment.

During the experiment you will have the chance to earn points, which will be converted into cash at the end of today’s session, using an exchange rate of 1 Guilder = £0.75.

In the experiment you are randomly matched with another participant – your partner. You will not learn the identity of the participant you are matched with, and vice versa your partner will never learn about your identity.

The experiment has a fixed and predetermined number of rounds.

You and your partner both receive an endowment of 10 Guilders. You then have to decide whether to reduce your partner's income or to leave it as it is. Reducing your partner's income will cost you 1 Guilder. By paying 1 Guilder, you can reduce your partner's income by 5 Guilders. Your partner answers some hypothetical questions but makes no decisions affecting your or his or her earnings.

If you choose to leave the other person's income unaltered, both of you will earn the 10 Guilders.

If you choose to reduce your partner's income you will earn 9 Guilders and your partner will earn 5 Guilders.

Some of these predetermined rounds will be marked with yellow instructions; this is instructions in yellow paper. In the rounds with yellow instructions it would be especially useful if you were to reduce your partner’s income if you have not done so already. You are entirely free not to reduce if you wish.

Instructions (RO1 - passive subjects)

Thank you for participating in this experiment. The experiment will last approximately 1 hour. If you read the following instructions carefully, you can, depending on your own decisions, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. It is not allowed to communicate with the other participants during the experiments. Should you have any questions, please ask us. If you violate this rule, we will have to dismiss you from the experiment and you will forfeit all payments.

You will be paid after the experiment. No other participant will know how much you earned. You will be paid £2 for showing up plus any additional earnings that you have in the experiment.

During the experiment you will have the chance to earn points, which will be converted into cash at the end of today's session, using an exchange rate of 1 Guilder = £0.75.

In the experiment you are randomly matched with another participant – your partner. You will not learn the identity of the participant you are matched with, and vice versa your partner will never learn about your identity.

The experiment has a fixed and predetermined number of rounds.

You and your partner both receive an endowment of 10 Guilders. Your partner has to decide whether to reduce your income or to leave it as it is. Reducing your income will cost him or her 1 Guilder. By paying 1 Guilder, he or she can reduce your income by 5 Guilders. You are asked to answer hypothetical questions where you simultaneously take the same decision but your decision does not count towards your or his/her earnings.

If he or she chooses to leave your income unaltered, both of you will earn the 10 Guilders.

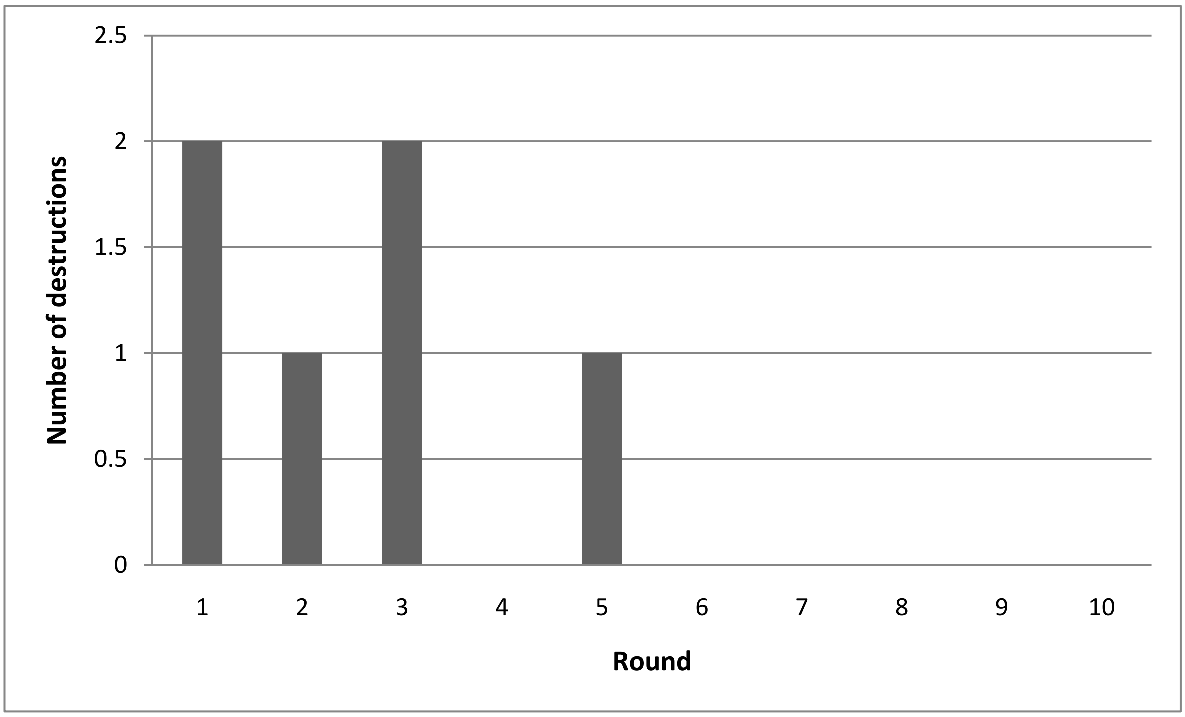
If he or she chooses to reduce your income he/she will earn 9 Guilders and you will earn 5 Guilders.

Some of these predetermined rounds will be marked with yellow instructions; this is instructions in yellow paper. In the rounds with yellow instructions it would be especially useful if you were to reduce your partner's income if you have not done so already. You are entirely free not to reduce if you wish.

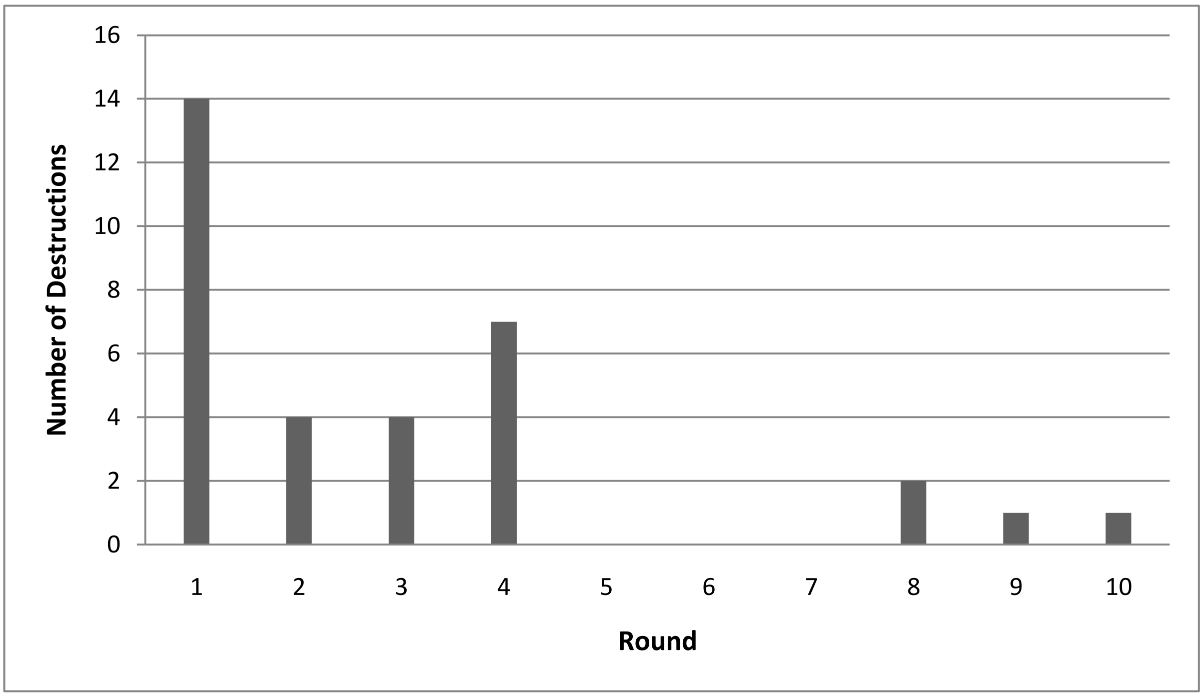
B. Number of Decisions to Destroy per Round by Treatment

Below we used 'number of destructions' as a shortcut for 'number of decisions to destroy'.

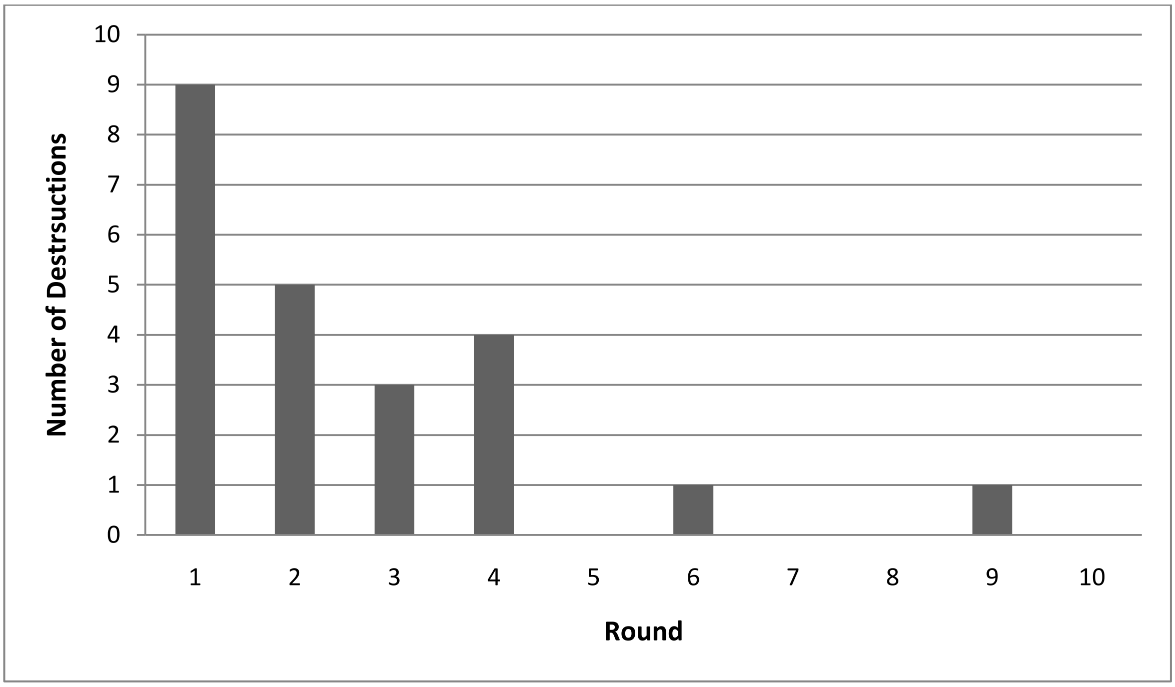
Number of Destructions per Round in the R Treatment



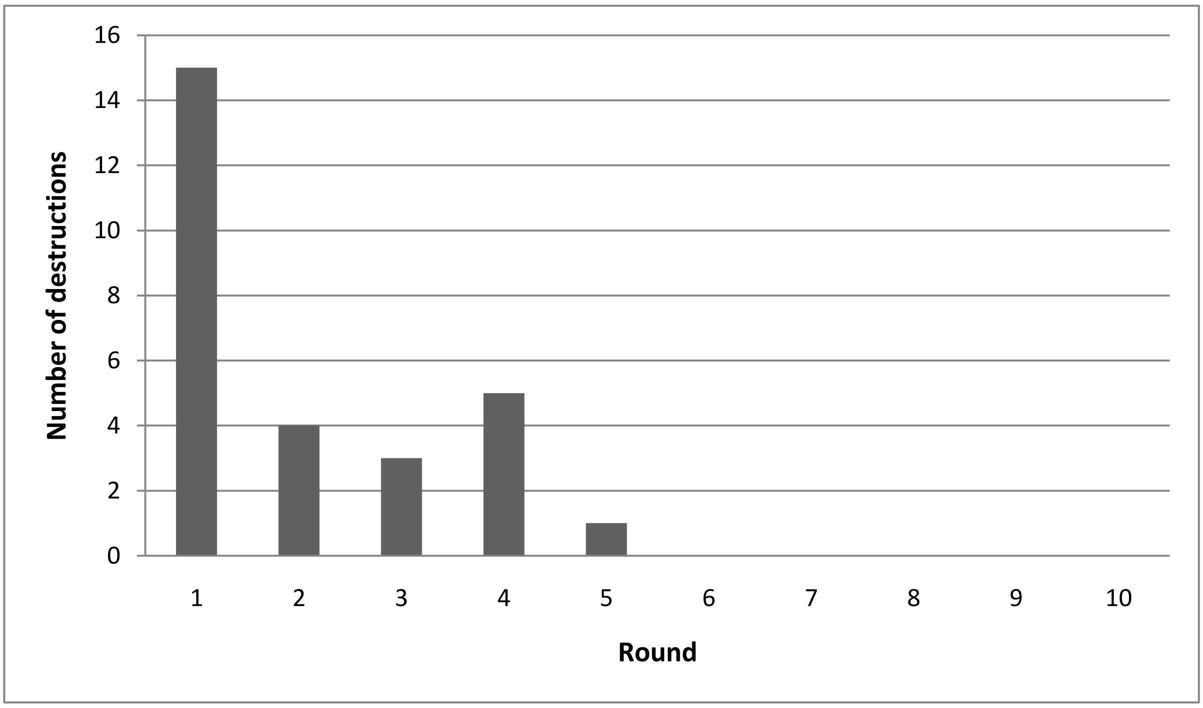
Number of Destructions per Round in the RO Treatment



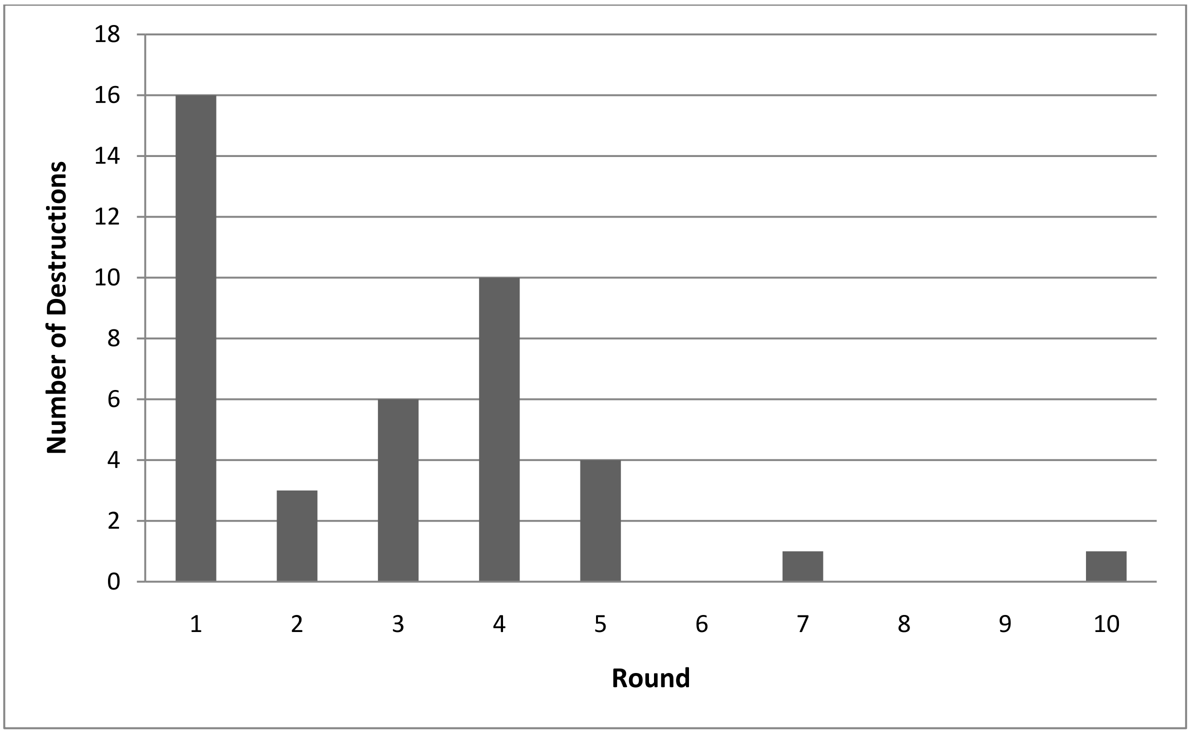
Number of Destructions per Round in the ROC Treatment



Number of Destructions per Round in the ROJ Treatment



Number of Destructions per Round in the RO1 Treatment



C. Round 1 Decisions

Table C1 provides information on first round destruction rates across treatments. the effect of the request to destroy is significantly larger even for the first round in ‘yellow round’ treatments for the RO, ROJ and RO1 treatments when these are compared to the O or R treatments (Fisher’s exact test $p < 0.05$, $p < 0.01$ and $p < 0.05$ respectively), noting that round 1 was a ‘yellow round’. There is no statistically significant difference in first round destruction rate between the ROC and the O and a weakly statistically significant difference between ROC and R treatments instead (Fisher’s exact test $p = 0.16$ and $p = 0.07$ respectively).

Regression C2 below provides Probit regressions on whether a destruction decision in round 1, bearing in mind that in treatments O and H there is not the possibility to delay destruction whereas this can take place in the other treatments. There is some mild evidence ($p < 0.1$) that economics students chose to destroy less in round 1, though they did not manage to keep this up overall (see Table 2 in main paper).

Table C1: Destruction rates in the first round

Treatment	First Round Destruction Rate (FRDR)	FRDR as a % of Overall Treatment Destruction Rate
O	7.5%	100%
H	15.0%	100%
R	5.0%	28%
RO	25.0%	42%
ROJ	37.5%	53%
RO1	25.0%	42%
ROC	16.0%	39%

	Regression C1			Regression C2		
	b	se	p	b	se	p
O	-0.780**	0.36	0.030	-0.765**	0.36	0.034
H	-0.399	0.31	0.193	-0.402	0.31	0.190
R	-1.001***	0.37	0.007	-0.994***	0.37	0.007
ROJ	0.361	0.28	0.192	0.364	0.28	0.190
RO1	0.033	0.29	0.910	0.036	0.29	0.904
ROC	-0.284	0.25	0.308	-0.286	0.25	0.305
British	-0.311	0.22	0.151	-0.299	0.22	0.171
Chinese	-0.261	0.28	0.354	-0.280	0.29	0.334
Gender	0.061	0.17	0.713	0.051	0.17	0.764
Age	-0.023	0.02	0.281	-0.220	0.02	0.287
Economics	-0.497*	0.27	0.065	-0.480*	0.27	0.073
Christian	-0.278	0.36	0.442	-0.261	0.36	0.471
Atheist	-0.361	0.35	0.302	-0.346	0.35	0.324
Muslim	-0.873	0.65	0.181	-0.876	0.65	0.181
Constant	-0.119	0.38	0.754	0.135	0.38	0.724
N	310			310		
Pseudo R-sqr	0.098			0.098		
Prob > X^2	0.011			0.016		

	Marginal Effects given Regression 1			Marginal Effects given Regression 2		
	b	se	p	b	se	p
O	0.074*	0.04	0.076	0.077*	0.04	0.067
H	0.139***	0.05	0.008	0.142***	0.05	0.007
R	0.185***	0.06	0.003	0.050	0.03	0.125
RO	0.611****	0.06	0.000	0.247****	0.06	0.000
ROJ	0.690****	0.07	0.000	0.370****	0.08	0.000
RO1	0.589****	0.08	0.000	0.258****	0.07	0.000
ROC	0.422****	0.07	0.000	0.168***	0.05	0.001

D. Fictional Destruction

D1. *Fictional Further Destruction*

Subjects who decided to reduce their partner's income were afterwards asked hypothetically if there would be willing to reduce their partner's income by an additional 5 ECU leading them to earn 0 ECU from this experiment. Note that fictional destruction could only take place in treatments with repeated interaction: the R, RO, ROJ, RO1 and ROC treatments.

Table D1 presents the average fictional destruction rates for each treatment. The middle column shows the percentage of subjects who where willing to destroy their partner's income even further according to their response on the hypothetical question. The last column shows the percentage of subjects who answered yes on the hypothetical question among all subjects who participated in the treatment. Figure 9 provides a visual representation.

Table D1: Fictional Destruction Rates

Treatment	Across subjects who have already destroyed	Across all subjects in treatment
R	83.0%	12.5%
RO	71.9%	42.3%
ROJ	96.4%	67.5%
RO1	65.2%	39.1%
ROC	78.3%	33.3%

Note that this data is of very limited value in across treatments comparisons, both because of its fictional nature, and because one cannot disambiguate what is due to treatment specific effects and what is due to sample selection of subjects who have already destroyed, and who are in different proportions in the different treatments. There is some evidence that, for subjects who were given a explicit reason to destroy (ROJ), the fictional further destruction is significantly larger than when no explicitly justification is provided (RO) (Fisher's exact test p = 0.01).

D2. *Fictional Destruction Rates of Passive Subjects in the RO1 Treatment*

Passive subjects in the RO1 treatment were asked to make fictional destruction choices, fully aware that their choices had no bearing on the actual game being played; if they said they would have destroyed, they were then asked a second time whether they would be willing to engage in further fictional destruction of another 5 pounds of the partner. Table D2 below compares the average destruction rate for the RO1 Treatment for both passive and active subjects, and for both the first time (real for active subjects, fictional for passive subjects) and for the second time (fictional for all). The patterns are clearly quite similar, with no statistical significance between groups for both the first and second destruction choice (Fisher's exact test p = 0.41 and p = 0.96 respectively).

Table D2: Fictional Destruction Rates in the RO1 Treatment

(as a proportion of all active subjects or of all passive subjects)

Treatment	Destruction choice (1st time)	Destruction choice (2nd time)
Active Subjects	60.0%	65.2%
Passive Subjects	55.0%	47.5%