

Distrust – The Hidden Cost of Control*

Armin Falk[†] Michael Kosfeld

June 2004

Abstract

We show experimentally that a principal's distrust in the voluntary performance of an agent has a negative impact on the agent's motivation to perform well. Before the agent chooses his performance, the principal in our experiment decides whether he wants to restrict the agents' choice set by implementing a minimum performance level for the agent. Since both parties have conflicting interests, restriction is optimal for the principal whenever the latter expects the agent to behave opportunistically. We find that most principals in our experiment do not restrict the agent's choice set but trust that the agent will perform well voluntarily. Principals who trust induce, on average, a higher performance and hence earn higher payoffs than principals who control. The reason is that most agents lower their performance as a response to the signal of distrust created by the principal's decision to limit their choice set. Our results shed new light on dysfunctional effects of explicit incentives as well as the puzzling incompleteness of many economic contracts.

JEL classification: C7, C9, M5

Keywords: trust, distrust, motivation, principal-agent relationship, incentives, control, incomplete contracts

*Financial support by the MacArthur Foundation (Network on the Evolution of Individual Preferences and Social Norms), the Ludwig Boltzmann Institute for the Analysis of Economic Growth, and the Swiss National Science Foundation (Project No 12-67751.02) is gratefully acknowledged.

[†]Corresponding author: University of Bonn and Institute for the Study of Labor, Schaumburg-Lippe Str. 7/9, D-53113 Bonn, Germany, E-mail: Falk@iza.org.

1 Introduction

Principal-agent relations are typically characterized by a conflict of interest. Therefore, principals often use control and incentive devices to eliminate agents' most opportunistic actions. Of course, if a principal believes that the agent will perform in the principal's interest and will not behave opportunistically, the introduction of control devices is superfluous. Controlling only makes sense if the principal expects the agent to behave opportunistically. Exactly for this reason, however, controlling signals distrust: by controlling the agent, the principal implicitly signals that he does not trust the agent to abstain from behaving opportunistically. This paper analyzes how the agent perceives the principal's controlling decision and how this affects the agent's behavior. We conducted an experiment in which a principal can decide either to trust or to control the agent, where controlling rules out the agent's most opportunistic actions. Our results show that the decision to control significantly reduces the agents' willingness to act in the interest of the principal. Explicit incentives backfire and performance is lower if the principal controls compared to if he trusts.

We analyze the interaction of motivation and control in a very simple and parsimonious set-up. In the game under study, an agent chooses a productive activity x which is costly to him but which increases the principal's payoff. The distinguishing feature of our experiment is the principal's decision. Before choosing x , the principal determines the agent's choice set. He can either leave the choice of x completely to the agent's discretion, in which case the lowest possible choice of x is zero. Alternatively, the principal can force the agent to choose at least a minimum level $\underline{x} > 0$. The definition of the agent's choice set can be interpreted as the degree of control that is implemented in the agent's work environment. For example, making it impossible for the agent to choose below \underline{x} is the equivalent of implementing different control or monitoring devices which restrain the agent from his most opportunistic choices. Not restricting the choice set, on the other hand, represents the absence of such control mechanisms. Alternatively, the restriction of the choice set can also be interpreted as the outcome of a corresponding employment contract. For example, if x represents the amount of working hours, \underline{x} captures a minimum presence requirement. Similarly, if x stands for the quality of a produced good or service, \underline{x} is the minimum quality the agent has to deliver.

Since x is costly to the agent, standard economic theory predicts that the agent will choose the lowest possible x , which is zero if the principal does not restrict the agent's choice set and $\underline{x} > 0$ if he does. Since the principal's payoff is increasing in x , he will therefore always be better off controlling the agent than not limiting the agent's choice set. However, if there are agents who are intrinsically motivated to perform in the principal's interest, controlling may actually decrease performance. A potential reason is that

agents do not like to be restricted and perceive control as a signal of distrust. Our main results, in fact, confirm this hypothesis. We find that a majority of the agents in our experiment reveals *control-averse behavior*, i.e., they choose a *lower* x if the principal restricts rather than trusts them. In fact, only a minority of the agents acts in line with the standard economic model, i.e., these agents always choose the lowest possible activity. As a consequence, principals earn significantly more if they trust their agents than if they control. Most of the principals in the experiment seem to have understood the cost of signaling distrust. The large majority decides not to restrict the agent’s choice set but to trust that the agent will perform well voluntarily. Principals who control have more pessimistic beliefs about the agent’s performance than principals who trust. As agents’ behavior roughly confirms both types of beliefs, our results nicely support the so-called “self-fulfilling prophecy of distrust” (Luhmann 1968).¹

We also conducted a control experiment to isolate the precise impact of control. The control experiment is a principal-agent game where the agent similarly chooses an activity x but is always forced to choose at least \underline{x} . The constraint \underline{x} is given exogenously in this treatment, i.e., the choice set of the agent is identical to the choice set in the main experiment when the principal decides to control. However, since the principal does not implement \underline{x} , distrust and control are not at issue in this treatment. If the principal’s controlling decision is behaviorally relevant to the agent, x should therefore be lower in the respective subgame of the main treatment than in the control treatment. This is in fact what we find.

Taken together, our results show that the use of explicit incentives entails “hidden costs” that should be considered when designing employment contracts and workplace environments. Elements in the labor contract that can be perceived as signals of distrust and control, such as minimum performance requirements, may harm more than they help. Similarly, characteristics of the workplace environment that limit freedom of choice and signal distrust, such as high levels of monitoring and surveillance, may be equally counterproductive.

In fact, the negative consequences of control offer a psychological rationale for the

¹The intuition of our results is neatly captured by an example reported by David Packard, one of the founders of the computer company Hewlett-Packard (HP). Packard notes in his memoirs: “In the late 1930s, when I was working for General Electric ..., the company was making a big thing of plant security. ... GE was especially zealous about guarding its tool and parts bins to make sure employees didn’t steal anything. Faced with this obvious display of distrust, many employees set out to prove it justified, walking off with tools and parts whenever they could. ... When HP got under way, the GE memories were still strong and I determined that our parts bins and storerooms should always be open. ... Keeping storerooms and parts bins open was advantageous to HP in two important ways. From a practical standpoint, the easy access to parts and tools helped product designers and others who wanted to work out new ideas at home or on weekends. A second reason, less tangible but important, is that the open bins and storerooms were a symbol of trust, a trust that is central to the way HP does business” (Packard 1995, pp 135).

incompleteness of many real-life economic contracts. In this sense, our paper offers a contribution to the literature on incomplete contracts. The degree of contractual incompleteness is an economic puzzle. Real-life contracts often omit important, verifiable obligations of the contracting parties (or only mention them vaguely), and measurable actions are often linked to verifiable information in a manner which is seemingly less than optimal. While bounded rationality on the part of the contracting parties might account for some of this incompleteness, the key question is why many contracts appear to be left *deliberately* incomplete. Several explanations have been offered to answer this question. In some situations, for example, incomplete contracts may be completed by renegotiation design to achieve first best outcomes, i.e., there is no need to write more complicated contracts (Aghion, Dewatripont, and Rey 1994, Nöldeke and Schmidt 1995). In other cases, incomplete contracts may actually be superior to more complete contracts. Allen and Gale (1992) and Spier (1992), e.g., argue that offering a more complete contract may lead the other party to draw negative inferences about the first party's type. Bernheim and Whinston (1998) show in a repeated game setting that if some obligations are non-contractible it may be better to leave other aspects unspecified, giving rise to so-called "strategic ambiguity" in the design of the contract. Our study provides a *behavioral* rationale for the deliberate incompleteness of many real-life contracts. We show that a large fraction of agents are averse to being controlled and consequently lower their performance if the principal implements a more complete contract. If the principal anticipates this effect, he may be better off choosing a less complete contract, leaving the agent substantial discretion and thereby signaling the principal's trust in the agent's non-opportunistic behavior.

Our paper also contributes to the recent literature dealing with the interaction of psychological and economic incentives (e.g., Frey 1997, Gneezy and Rustichini 2000a,b, Frey and Jegen 2001, Gneezy 2004). The analysis differs, however, in several dimensions with respect to previous studies that have demonstrated dysfunctional effects of explicit incentives.

First, principals in our experiment have a new and yet unexplored decision possibility. Rather than specifying punishments and rewards, a principal in our game can determine the choice set of the agent by fixing a positive minimum performance requirement. These possibilities are common in many real-life labor relations (regulated working times, high-control working environments, minimum output/quality, etc.) but, to the best of our knowledge, have not yet been explored in a principal-agent framework. In addition, our experimental design carefully separates control and trust from payoff-driven reciprocity and gift exchange. In contrast to previous experiments (Fehr and Gächter 2002; Fehr and Rockenbach 2003; Fehr and List 2004; Fehr, Klein, and Schmidt 2004) it does not

pay for the principal to trust in our model because trusting results in a higher payoff to the agent which is reciprocated. Instead, trust only pays in our experiment because some agents are intrinsically trustworthy and react negatively to signals of distrust.² A further distinction of our paper is that we are able to obtain detailed individual-level information on the agent's behavior rather than aggregate results alone. In particular, we can distinguish between three main behavioral types: selfish, inequity-averse, and control-averse agents. Finally, our study highlights a new mechanism for possible dysfunctional effects of incentives, namely distrust. Previous studies have shown that incentives may undermine motivation because they provide new information regarding the importance or the cost of the task (Gneezy and Rustichini 2000a, Benabou and Tirole 2003), because they insult the agent (Gneezy and Rustichini 2000b), or because they are in conflict with social norms of fairness and cooperation (Fehr and Gächter 2002, Fehr and Rockenbach 2003, Fehr and List 2004, see also the discussion in Gneezy 2004).

Since our results provide empirical evidence for negative effects of control, our paper is also closely related to Frey (1993) who proposes a theoretical framework for analyzing the (possibly negative) effects of monitoring on agents' work effort.³ The notion of control aversion is consistent with results from social psychology showing that individuals are often less motivated to do a certain task when they are monitored either by a physically present surveillant (Pitman et al. 1980) or by means of a video camera (Lepper and Green 1975, Plant and Ryan 1985). Enzle and Anderson (1993) demonstrate that the negative motivational effect is particularly strong if the surveillance is seen to reflect controlling purposes.

The paper is organized as follows. Details of the experimental design are explained in the following section. Section 3 discusses the behavioral predictions. Results of the experiment are presented in Section 4. Section 5 provides a short discussion and concludes.

2 Experimental design

Our design philosophy was to set up an experimental game that allows studying the potential interaction between control and motivation in a parsimonious way. We therefore implemented the following two-stage principal-agent game. The agent chooses a productive activity x , which is costly to him but beneficial for the principal. The cost for the

²This difference can, perhaps, best be seen from the fact that none of the present models of reciprocity and social preferences is able to predict the results of our experiment (see Section 3).

³Two recent papers that argue in a similar direction are Sliwka (2003) and Ichino and Mühlheusser (2004). Barkema (1995) is the only empirical study we know of that analyzes how monitoring affects workers' effort, where the latter is measured by the number of working hours. The study uses data from 116 executives of medium-sized Dutch firms. While the author finds that in some cases higher monitoring actually correlates with fewer working hours, the causal relationship between the two variables is unclear.

agent is $c(x) = x$, while the benefit for the principal is $2x$, i.e., the marginal cost of providing the productive activity is always smaller than the marginal benefit. The agent has an initial endowment of 120 while the endowment of the principal is 0. The payoff functions are thus given by

$$\pi_p = 2x$$

for the principal and

$$\pi_a = 120 - x$$

for the agent. The crucial feature of our design is the principal's choice. Before the agent decides on x , the principal determines the agent's choice set. He can either restrict the agent's choice set, in which case the latter can choose any integer value $x \in \{10, 11, \dots, 120\}$, or he can leave the choice set unrestricted to $x \in \{0, 1, \dots, 120\}$. Thus, the principal can (partly) control the decision environment of the agent, thereby guaranteeing a minimal payoff of 20, or he can leave the decision completely up to the agent trusting that the latter will not choose x below 10.

We made use of the strategy method to elicit the agents' choices in the experiment. Before knowing the actual decision of their principal, the agents had to decide on x for both possible cases. On a computer screen they were asked to choose $x \in \{10, 11, \dots, 120\}$ in case their principal forces them to choose at least 10, and to choose $x \in \{0, 1, \dots, 120\}$ in case their principal does not restrict their choice set. We used the strategy method to gain direct information about individual types which will be discussed in detail below. The game was played one-shot. It was framed in a neutral manner. In particular, we did not speak of trust or distrust in the instructions.

We also ran a control experiment. In this treatment, the principal and the agent just played the subgame of the main treatment that follows the restriction choice of the principal. As before the agent is endowed with 120 points. He chooses a productive activity $x \in \{10, 11, \dots, 120\}$. Payoffs are $120 - x$ for the agent and $2x$ for the principal. The principal makes no decision.

The experiment was computerized using the software "z-Tree" (Fischbacher 1999) to run the experiment. Subjects were randomly allocated a role as principal or as agent upon arrival at the lab. Subjects were students from the University of Zurich and the Swiss Federal Institute of Technology in Zurich. No subject participated both in the main and the control treatment. In the main treatment a total of 144 subjects participated (72 principals and 72 agents). A session lasted, on average, 40 minutes. On average, a subject earned CHF 24.10 (about \$19) in the experiment. 72 subjects participated (36 principals and 36 agents) in the control treatment. It lasted about 40 minutes and subjects earned CHF 24.90 on average.

3 Behavioral predictions

The behavioral outcome of the main treatment depends on the assumption concerning players' preferences. Let us consider the standard case first. If agents are selfish, i.e., only interested in maximizing their own income, they choose the minimum x . This implies that they choose $x = 10$ if they are restricted and $x = 0$ otherwise. In consequence, principals who want to maximize their payoff should always control the agent's choice set. According to this benchmark solution, agents are opportunistic and therefore it inevitably pays to rule out their most opportunistic choices.

There is ample evidence, however, that many people do not act in a purely selfish manner but are endowed with social preferences (for an overview of the experimental literature see, e.g., Fehr and Gächter 2000 and Camerer 2003). In the dictator game, e.g., proposers often give positive amounts of money (Forsythe *et al.* 1994). These findings were recently explained in terms of subjects' concern for equity and efficiency (Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Charness and Rabin 2002). In addition to their preference for material payoffs, subjects dislike inequitable outcomes, i.e., they are willing to sacrifice money in order to reach more equitable outcomes. Similarly, the efficiency of the outcomes might also play a role. In light of the experimental evidence and according to social-preference theories, a substantial fraction of agents in our principal-agent game are therefore expected to choose an activity x that is strictly larger than the respective minimum. However, if the choice of the agent is larger than 10 (i.e., the agent has a relatively strong preference for equitable outcomes), inequity-aversion implies that the agent's choice is in fact *independent* of whether a principal controls the agent or not. This follows because inequity-aversion preferences are based on payoff distributions and the constraint $x \geq 10$ is not binding. Only weakly inequity-averse agents who choose $0 < x < 10$ if not controlled, will choose a higher activity (namely, $x = 10$) if controlled. Taken together, as it holds for the standard economic case, a payoff maximizing principal cannot lose anything (but can possibly gain something) from controlling inequity-averse agents and hence should always restrict the agent's choice set.⁴ Note further that inequity aversion predicts no difference in behavior between the control treatment and the subgame of the main treatment that follows the restriction decision.

⁴Formally, according to the model of Fehr and Schmidt (1999) inequity-averse agents choose either $x = 0$ if $\beta < \frac{1}{3}$ or $x = 40$ if $\beta > \frac{1}{3}$. In the first case it pays to control, in the second case the agent's choice is unaffected. Choices other than 0 and 40 can be explained by non-linear versions of the model. Other theories of social preferences exist that are based not only on payoff distributions but which take also players' intentions into account (Rabin 1993, Falk and Fischbacher 1998, Dufwenberg and Kirchsteiger 2004). None of these theories, however, predict that the principal should refrain from controlling the agent's choice set. The reason is that the principal's decision not to control the agent can only be perceived as a "kind action" if the agent actually intends to exploit the principal's trust by choosing a low performance, in which case the principal, of course, has no incentive to trust the agent.

Now suppose that agents do not like to be controlled. We call someone *control averse* if he dislikes being controlled and reduces performance in response to being controlled. A potential reason why we might see this behavioral response is that controlling signals distrust. The principal will decide to control the choice set of his agent if he believes that the agent will perform poorly otherwise, i.e., will choose $x < 10$. Thus, if the principal controls the agent, the message conveyed to the agent is that he is not trusted to perform well. A control-averse agent reacts to this signal by choosing a lower performance than he would have chosen in case the principal had not controlled the agent. If faced with control-averse agents, the optimal strategy for the principal is thus the opposite of what was the optimal strategy as described above: a principal is better off trusting and not controlling the agent. Control aversion further implies that x -choices are higher in the control treatment than in the corresponding subgame of the main treatment. The reason is that while the strategy set and the payoff consequences are identical, the principal does not choose the constraint \underline{x} , but it is imposed exogenously in the control treatment. Hence, the constraint cannot signal distrust.

The quantitative importance of the three behavioral motives is, of course, an empirical question. If the fraction of selfish and inequity-averse agents is sufficiently high, it will pay for the principal to control the agent's choice set and eliminate his most opportunistic choices. If, on the other hand, there are relatively more control-averse agents, a principal will be better off trusting the agent. In the next section, we will not only answer this question on an aggregate level. Since we employ the strategy method to elicit the agents' choices, we can directly identify all three motivational types and determine their relative frequencies.

4 Results

In the results section we first present the agents' behavior on the aggregate and on the individual level. We then discuss the results of the control experiment. Finally, we turn to the behavior of the principals.

4.1 Agents

Our first result shows that, on average, restricting the agent's choice set has an adverse effect on the agent's motivation to choose a high x .

Result 1 *Agents choose significantly higher values of x if the principal does not control than if he does so.*

Support for Result 1 comes from Table 1 and Figure 1. Table 1 shows the average and median choices as well as the standard deviations of x elicited from the strategy method, i.e., if the principal does and does not control the agent’s choice set. This table conveys a clear message. Both average and median values of x are higher in the no-control case compared to the control case. On average, agents choose 31 percent higher values of x if they are not controlled by their principal and the median is twice as high. These differences are statistically significant at any conventional level (Wilcoxon signed rank test, $p < .001$).⁵ Thus, both in contrast to the standard economic model and to the model of inequity aversion, agents not only choose above the minimum, they also choose more if the required minimum is lower.

Table 1: Agents’ choices dependent on the principal’s decision

	Principal controls	Principal does not control
Average	17.5	23.0
Median	10	20
Standard deviation	13.57	17.97
Number of observations	72	72

Figure 1 shows how the x -choices are cumulatively distributed given the decision of the principal.⁶ Obviously, the distribution of x is very different under the two conditions. Most notably, if the principal controls the agent, the by far most frequent choice of x is 10 (dark points). When being forced to choose at least 10, 53 percent of the agents choose exactly the minimum $x = 10$ and only about 20 percent choose $x > 20$. In sharp contrast, only 32 percent choose $x \leq 10$ in the no restriction case and more than 47 percent of the agents choose $x > 20$. While only 7 percent of the agents choose the payoff equalizing level of $x = 40$ if they are controlled by the principal, 24 percent do so if they are not controlled. Taken together, Figure 1 shows that we see relatively high x -choices from the

⁵Note that the Wilcoxon test assumes that the distribution of the differences between the values within each pair of observations is symmetrical. Since the controlling decision of the principal changes the agent’s strategy space, this assumption is likely to be violated. Note, however, that this does not question the interpretation of the results. First, since agents are forced to choose a higher x if controlled than if not controlled, results are – if anything – biased *against* the control-aversion hypothesis. Second, we can modify the distribution of x -choices if not controlled such that any observation $x < 10$ is set equal to 10. Under the assumption that control has no behavioral impact, the resulting modified distribution should be the same as the one generated if controlled. Given the results shown so far it is, however, not surprising that they are in fact different from each other ($p < .0001$). The average (median) of the modified distribution is 25 (20).

⁶Figure 1 shows all observations for $x \leq 50$. In both conditions there was one x above 50. In the control case this was $x = 95$, in the no-control case it was $x = 105$. These two cases are summarized in Figure 1 as $x > 50$.

agents if they are in an uncontrolled decision environment, whereas the mass of x -choices is centered at 10 if the principal restricts the agent's choice set.

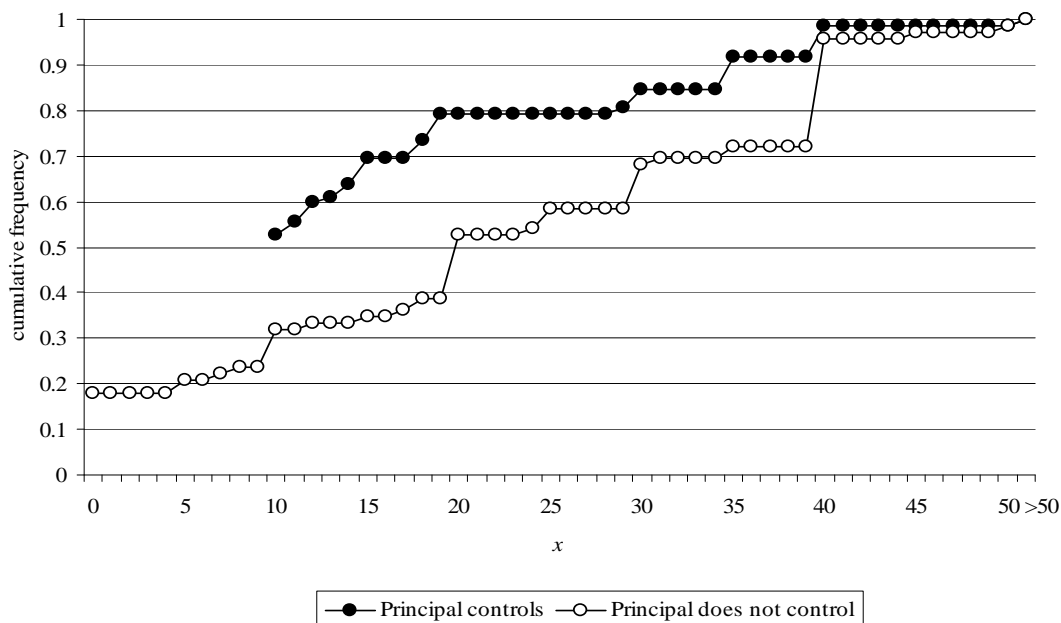


Figure 1: Cumulative distribution of agents' choices dependent on the principal's decision

The aggregated results shown in Table 1 and Figure 1 provide a first indication that there are indeed agents who reveal control aversion, i.e., who choose lower performance levels if controlled than otherwise. However, there must also be other types of agents given that 32 percent choose $x \leq 10$ in the no-control case. Since agents had to state their x -choices for both possible cases, we are in the position to determine the relative frequency of the different types of behavioral motivations.

Result 2 *There is a strong heterogeneity in agents' types. We see selfish, inequity-averse and control-averse agents, the last group being the majority.*

Support for Result 2 comes from Figure 2 and Table 2. Figure 2 plots the choice of x for each of the 72 agents on an individual basis if the principal controls the agent minus the choice of x if the principal does not control him. If all agents were selfishly motivated, this difference should be positive and close to 10 for every agent. Figure 2 instead shows that there are three main types of behavioral patterns. For the first group of agents to the very left, the effect of controlling the agent's choice set is in fact positive as hypothesized by the standard economic model. 25 percent of the agents belong to this group. For a second group of agents, the decision of the principal seems to be irrelevant as they choose the same x in both cases. This group consists of 18 percent of all agents. Finally, control

has a negative effect on performance for 57 percent of the agents. They choose a lower x if their choice set is restricted than otherwise.

Figure 2 summarizes the costs and benefits of the principal’s decision to control the agent. If agents were only of the first two behavioral types, it would always pay to control. Either the agent is opportunistic and simply chooses the minimum performance level or he does not care about possible signals of distrust. Given the large fraction of agents who show a negative behavioral reaction to control, however, it is obvious that the “hidden costs” of controlling the agent’s choice set are nonzero and even outweigh the benefits.

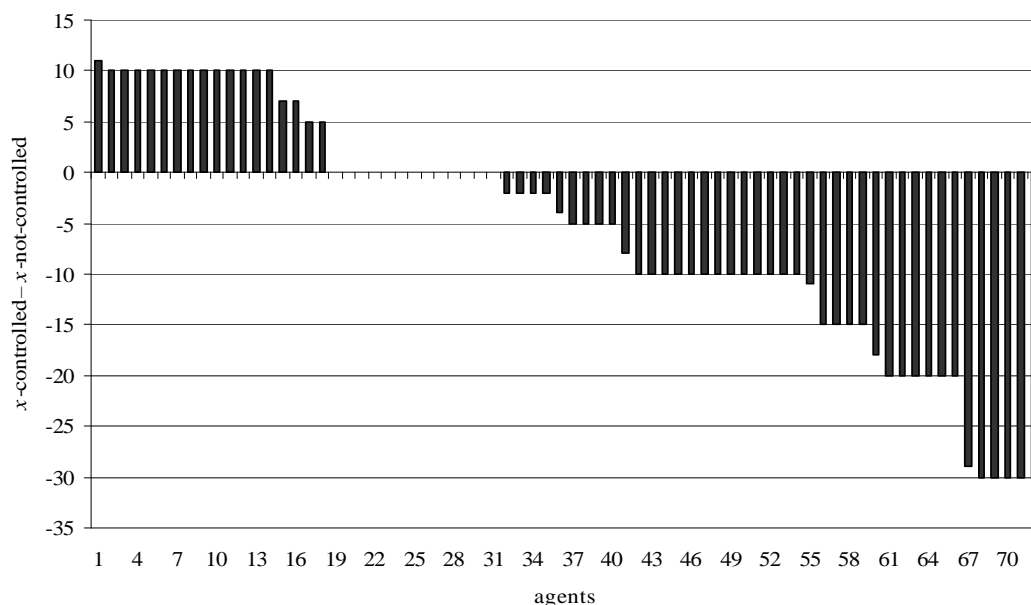


Figure 2: Costs and benefits of control: individual differences

Figure 2 shows for each agent whether the performance response to being controlled by the principal is positive, neutral, or negative. A closer inspection of the data allows an even more detailed classification (see Table 2). In light of our discussion in the previous section, we classify an agent as selfish if he chooses 0 or 10 dependent on being controlled or not. 18 percent of the agents are of this type. A subject is called inequity averse if he chooses more than zero when being free to choose and not less when being controlled. Among these agents, who represent 21 percent of the cases, most agents choose in fact the same x in both situations most often either 10 or 40. The largest group of agents, however, consists of control-averse individuals who show a negative response to being restricted in their choice set. 57 percent reveal this behavioral pattern. They are responsible for why, on average, it does not pay for the principal to control. Their mean (median) choice of x if controlled is 18.7 (12) while it is 32.3 (30) if they can decide freely. Finally, there are three subjects, who do not fall in any of the three categories. These agents are not selfish

according to our definition but choose a higher x if they are controlled than otherwise.⁷

Table 2: Distribution of agents' type

	Selfish ⁸	Inequity averse	Control averse	Other
Number	13	15	41	3
Percent	18	21	57	4
Mean x if controlled	10.1	21.0	18.7	16.3
Mean x if not controlled	0.0	20.3	32.3	8.3

4.2 Control treatment

So far we have shown that control has an adverse effect on agents' performance. In principle, this effect could be due to the mere difference in the agent's strategy space, i.e., to the difference in the support of x between the two conditions. To isolate the effect of the principal's control decision, we therefore conducted an additional experiment where the support of x is the same as in the control condition but control on the part of the principal is absent. This is achieved by a game where the constraint $x \geq 10$ is given exogenously to the agent. In this game, the agent's strategy space is identical to the respective subgame in the main treatment. The only difference is that the agent is controlled by the principal in the main treatment, while there is no control-decision in the control treatment. If controlling has a negative impact on agents' motivation, the x -choices in the control treatment should exceed those in the main treatment. Our next result shows that this is indeed what we find.

Result 3 *Agents' behavior is sensitive to the principals' decision to control: x is significantly higher in the control treatment than in the main treatment.*

Support for Result 3 comes from Figure 3. It shows the cumulative distribution of agents' x -choices in the main treatment if the principal decides to control together with the cumulative distribution of agents' x -choices in the control treatment.⁹

The data unambiguously show the adverse motivational effect of the principal's decision to control. Agent's choices of x are significantly lower in the main treatment than in the control treatment. The average (median) choice of an agent is 17.5 (10) if the

⁷In Appendix 1 we report all choices of the agents together with how they are classified.

⁸One of the subjects classified as selfish actually chose $x = 11$ if controlled and $x = 0$ if not. This is why the mean x if controlled is 10.1 and not 10.

⁹Again, Figure 3 shows all observations for $x \leq 50$. There were two x -choices above 50 in the control experiment (70 and 119). These are summarized in Figure 3 as $x > 50$.

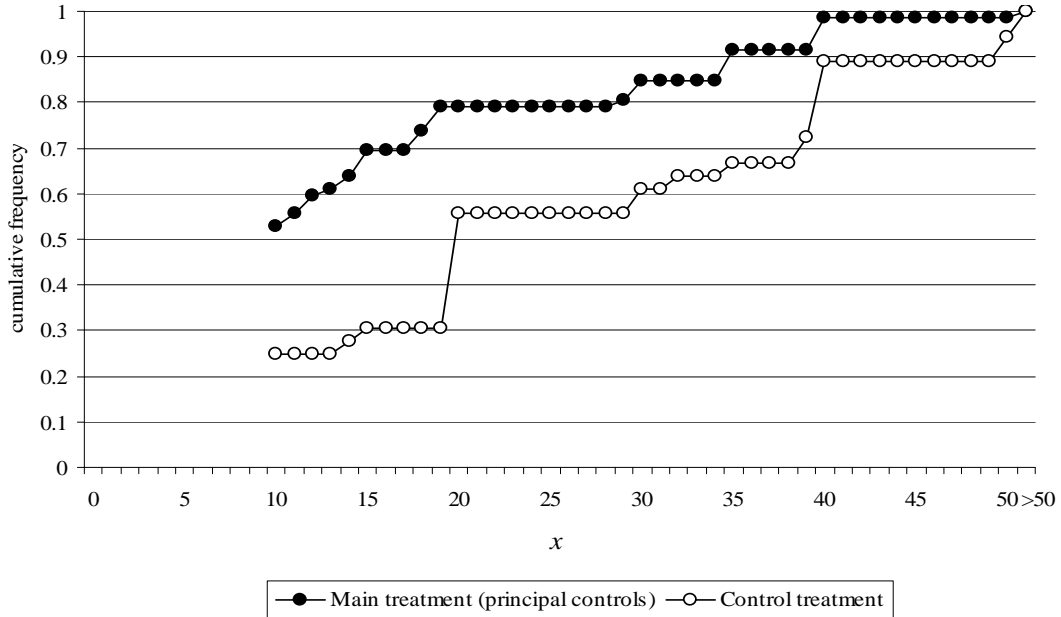


Figure 3: Cumulative distribution of x in the control treatment and the corresponding subgame of the main treatment

principal restricts the agent’s choice set compared to 28.7 (20) if the restriction is imposed exogenously. These differences are statistically significant at any conventional level (Mann-Whitney test, $p < .001$). While 53 percent of the agents choose the minimum performance level if the principal forces them to choose at least 10, only 25 percent choose the minimum $x = 10$ in the control treatment. On the other hand, about 45 percent choose $x > 20$ in the control treatment whereas only about 20 percent do so in the main treatment.

4.3 Principals

We now turn to the principals. Given the average responses of the agents, it is clearly not optimal for the principals to control the agent. However, if they do not control, they risk getting nothing while they can guarantee at least an income of 20 if they do. The principals should therefore refrain from controlling only if they correctly anticipate the adverse effects of doing so. Our next result shows that most of the principals seem to have understood this.

Result 4 *The large majority of the principals chooses not to control the agent.*

Result 4 is supported by the fact that 70.8 percent of the principals decide not to restrict the agent’s choice set while 29.2 percent do. This difference is statistically significant at any conventional level (Binomial test, $p < .001$). The data suggest that principals

anticipate possible adverse effects of controlling their agent because they voluntarily give up the option to guarantee 20 points for themselves.

In order to better understand the principals' decisions, we also elicited their beliefs about the agent's choice of x . Right after principals had made their decision in the experiment, they were asked to state their expectation about the agent's choice of x as well as how certain they are about this expectation.¹⁰ The main finding is summarized in our next result.

Result 5 *Principals who control have lower expectations about x than principals who don't.*

The mean (median) expectation about x for those who control is 19.4 (10) while it is 25.7 (25) for those who don't. The differences in beliefs are statistically significant (Mann-Whitney test, $p = 0.0046$).¹¹ In addition, principals who control are more certain about the expected performance of the agent than principals who don't. The mean (median) value is 4.8 (5) if the principal controls and 3.6 (3) otherwise. Figure 4 shows the cumulative distribution of the principals' beliefs with regard to x .¹² A comparison of the agents' actual choices displayed in Figure 1 with the principals' expectations in Figure 4 reveals that expectations were indeed quite accurate. Note in particular that 57 percent of the principals who force their agent to choose at least 10 expect to get $x = 10$. The actual number is 53 percent, which is very close to this expectation. In the no-control case, principals are somewhat too optimistic. For example, they expect the equal split in 39 percent of the cases but agents choose $x = 40$ only in 24 percent of the cases. However, principals' beliefs and agents' choices are not statistically different from each other. A Mann-Whitney test does not reject the hypothesis that beliefs and actions are the same both if principals do not control ($p = 0.2025$) and if principals do control ($p = 0.7603$).

The beliefs of the principals and the actual performance of the agents in our experiment illustrate what has been called the "self-fulfilling prophecy of distrust" (Luhmann 1968). Principals, who have rather pessimistic beliefs and hence choose to restrict the agent's choice set, experience that their beliefs are indeed confirmed by the relatively low average performance of their agents. On the other hand, principals who have optimistic beliefs and trust their agents, see their beliefs roughly confirmed as well.

In consequence, different and reinforcing "firm cultures" may emerge. Managers in low trust firms, on the one hand, have little trust in their employees and predominantly

¹⁰Beliefs were not paid. Certainty ranged from 1 (very uncertain) to 7 (very certain).

¹¹If we modify the distribution of beliefs in the no-control condition such that all beliefs smaller than 10 are set equal to 10, the p -value of the Mann-Whitney test is .0009. See our discussion in footnote 5.

¹²Figure 4 shows all observations for $x \leq 50$. In the restriction condition one principal stated an expected x above 50. This is displayed in Figure 4 as $x > 50$.

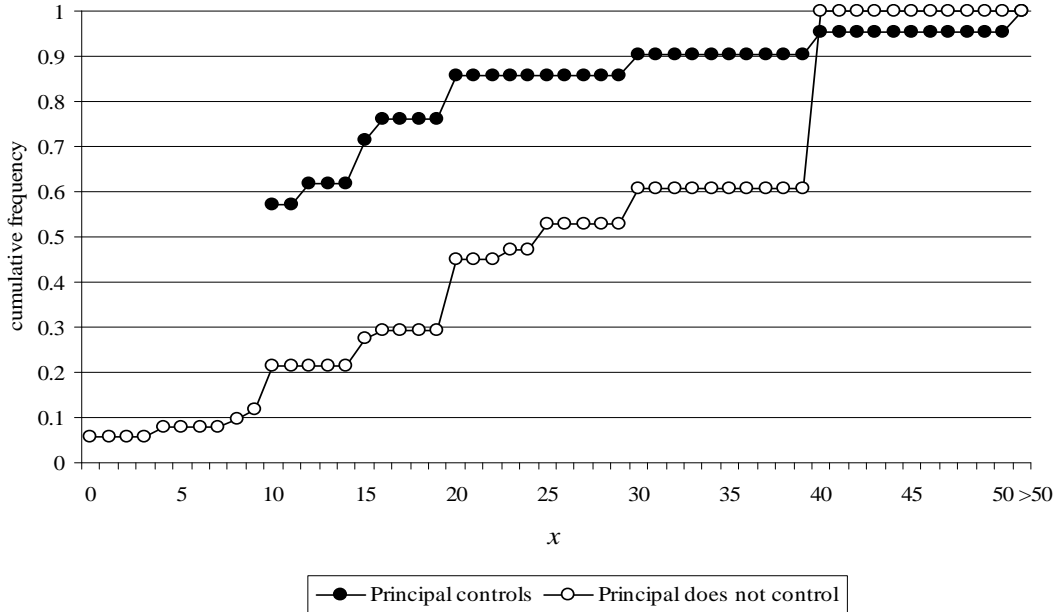


Figure 4: Cumulative distribution of principals' belief about x , dependent on their decision to control or not to control

rely on control. These managers will not be surprised to see performance at the minimum, confirming their pessimistic beliefs. The locked storeroom policy of GE in the 1930's is a perfect example for such firm culture (see footnote 1). High trust firms, on the other hand, are governed by the expectation of mutual trust. Employees are trusted in these firms and given responsibility. This "empowerment" of agents actually produces non-minimal performance, substantiating the initial beliefs held by the managers.¹³

5 Discussion

In this paper, we analyze the interaction of control and motivation in a principal-agent relationship. We introduce a simple game, which allows studying the potential impact of control in a parsimonious way. Our results show that a majority of the agents exhibits control-averse behavior, i.e., they are less motivated to perform well if the principal forces them to provide a minimum level of performance. On average, a principal in our experiment earns 31 percent less if he controls the agent than if he does not control. These costs are absent in any standard economic model assuming either selfish or inequity-averse behavior on the part of the agent. Not surprisingly, therefore, the

¹³The stability of these prototypical firm cultures is probably different. Low trust firms will encounter deviations between expectations and actual behavior to a lesser extent than high trust firms. Remember that when the principals did not control, they were somewhat too optimistic (Figure 4). This implies that some principals who have been disappointed may adjust their beliefs and choose to control in the future. It would be interesting to study the stability of firm cultures in a repeated interaction framework.

subgame-perfect-equilibrium prediction of the standard model yields the complete opposite of what the large majority of the principals in our experiment actually does. While the standard model predicts that all principals restrict the agent’s choice set, 71 percent of the principals decide not to control but to trust instead.

We believe that the main mechanism behind control aversion is the principal’s signal of distrust in the voluntary performance of the agent that is generated by his decision to restrict the agent’s choice set. Our hypothesis is consistent with the results from the control treatment which show that agents’ behavior is a negative reaction to the principal’s decision. To shed further light on agents’ behavioral motive, we also asked subjects after the experiment whether the restriction by the principal is indeed perceived as a signal of distrust in the sense that the agent is not trusted to choose a positive x . Subjects could express their degree of agreement to the question on a five-level scale ranging from “do not agree at all” to “fully agree”. Subjects’ answers are as follows. If we take all subjects who participated in the role of an agent we find that 86 percent either agree or fully agree that the principal’s decision to restrict their choice set does signal distrust. Only 14 percent do not agree. Thus, even those agents who do not behave in a control-averse manner in the experiment agree that the restriction decision signals distrust. Among those agents who have been classified as control-averse even 93 percent (i.e., 38 out of 41 subjects) either agree or fully agree that the principal’s control decision signals distrust. We are aware of the potential difficulties with questionnaire data and therefore believe that subjects’ answers should be handled with care. In principle, we cannot rule out that subjects’ responses are biased simply because we mention the word distrust. Bearing this in mind, however, we still think that the large agreement of the agents (in particular, of the control-averse agents) seems remarkable and is therefore worth reporting.¹⁴

The main message of our paper is that explicit incentives and control entail “hidden costs” which should be taken seriously. The message is *not*, however, that it is always better for principals to trust than to control. The level of the hidden costs and the benefit of controlling depend on various factors. First, they depend on the composition of agents’ types. When facing rather opportunistic agents with a low intrinsic motivation to perform in the interest of the principal, controlling generates only minor costs and trusting is likely

¹⁴We can think of two alternative explanations for agents’ behavior apart from distrust. Each of these explanations, however, comes with some problems. First, the restriction decision of the principal might not signal the principal’s belief about the agent but rather the principal’s type. The idea is that a principal considers what he would do if he was in the situation of the agent and then reacts optimally to his own type of behavior. The problem with this explanation is that it does not allow for altruistic principals who are afraid of being cheated by a selfish agent and therefore choose to control. Next, agents’ behavior could be driven by a general motive to signal generosity, which becomes less important (or more costly) if agents have to give at least \underline{x} anyway. In this case, however, choices in the control treatment should be the same as in the main treatment because it does not matter where the lower constraint comes from.

to be suboptimal. Second, the strength of the explicit incentives is important. To illustrate, imagine the principals could have chosen, e.g., $\underline{x} = 20$ or $\underline{x} = 40$ in our experiment. In these cases, the benefits of controlling would have been substantially higher and controlling would probably have resulted in higher expected performance levels than trusting. In this sense, the relation between agents' performance and the strength of incentives is probably non-monotonic: if the principal only has weak incentives at his disposal it may be better to trust since controlling demotivates the intrinsically motivated agents but only increases the performance of opportunistic agents marginally. As incentives get stronger, however, the disciplining effect will eventually dominate the demotivating effect (on this point see also Gneezy 2004). Finally, the hidden costs of explicit incentives are very likely to depend on the type and the framing of incentives. For example, rewarding good results instead of punishing bad ones may be perceived very differently (cf. Fehr and Gächter 2002). While the former may actually reinforce motivation, the latter conveys a negative message, dampening work morale. Faced with agents who are very heterogeneous, the ultimate management goal is to develop incentive contracts that discipline the opportunistic agents without demotivating the intrinsically motivated ones.

References

- Aghion, P., Dewatripont, M., and Rey, P. (1994) “Renegotiation design with unverifiable information,” *Econometrica* 62, 257-282.
- Allen, F. and Gale, D. (1992) “Measurement distortion and missing contingencies in optimal contracts,” *Economic Theory* 2, 1-26.
- Barkema, H.G. (1995) “Do top managers work harder when they are monitored?,” *Kyklos* 48, 19-42.
- Benabou, R. and Tirole, J. (2003) “Intrinsic and extrinsic motivation,” *Review of Economic Studies* 70, 489-520.
- Bernheim, B.D. and Whinston, M.D. (1998) “Incomplete contracts and strategic ambiguity,” *American Economic Review* 88, 902-932.
- Bolton, G.E. and Ockenfels, A. (2000) “A theory of equity, reciprocity, and competition,” *American Economic Review* 100, 166-193.
- Camerer, C.F. (2003) *Behavioral Game Theory*, Princeton University Press, Princeton.
- Charness, G. and Rabin, M. (2002) “Understanding social preferences with simple tests,” *Quarterly Journal of Economics* 117, 817-169.
- Dufwenberg, M. and Kirchsteiger, G. (2004) “A theory of sequential reciprocity,” *Games and Economic Behavior* 47, 268-298.
- Enzle, M.E. and Anderson, S.C. (1993) “Surveillant intentions and intrinsic motivation,” *Journal of Personality and Social Psychology* 64, 257-266.
- Falk, A. and Fischbacher, U. (1998) “A theory of reciprocity,” IEW Working Paper 6, University of Zurich.
- Fehr, E. and Gächter, S. (2000) “Fairness and retaliation: the economics of reciprocity,” *Journal of Economic Perspectives* 14, 159-181.
- 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., Klein, A., and Schmidt, K.M. (2004) “Fairness, incentives, and contractual incompleteness,” mimeo, University of Zurich.

- Fehr, E. and List, J.A. (2004) "The hidden costs and rewards of incentives," *Journal of the European Economic Association* forthcoming.
- Fehr, E. and Rockenbach, B. (2003) "Detrimental effects of sanctions on human altruism," *Nature* 422, 13 March 2003, 137-140.
- Fehr E. and Schmidt, K.M. (1999) "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics* 114, 817-868.
- Fischbacher, U. (1999) "z-Tree: A toolbox for readymade economic experiments," IEW Working Paper 21, University of Zurich.
- Frey, B.S. (1993) "Does monitoring increase work effort? The rivalry with trust and loyalty," *Economic Inquiry* 31, 663-670.
- Frey, B.S. (1997) *Not just for the money*, Edward Elgar, Cheltenham.
- Frey, B.S. and Jegen, R. (2001) "Motivation crowding theory," *Journal of Economic Surveys* 15, 589-611.
- Forsythe, R., Horowitz, J.L., Savin, N.E., and Sefton, M. (1994) "Fairness in simple bargaining experiments," *Games and Economic Behavior* 6, 347-369.
- Gneezy, U. (2004) "The W effect of incentives", mimeo, The University of Chicago Graduate School of Business.
- Gneezy, U. and Rustichini, A. (2000a) "A fine is a price," *Journal of Legal Studies* 29, 1-17.
- Gneezy, U. and Rustichini, A. (2000b) "Pay enough or don't pay at all," *Quarterly Journal of Economics* 115 (2), 791-810.
- Ichino, A. and Mühlheusser, G. (2004) "How often should you open the door? Optimal monitoring to screen heterogeneous agents," IZA Discussion Paper 987, Bonn.
- Lepper, M.R. and Greene, D. (1975) "Turning play into work: effects of adult surveillance and extrinsic rewards on children's intrinsic motivation," *Journal of Personality and Social Psychology* 31, 479-486.
- Luhmann, N. (1968) *Vertrauen*, 4th Edition 2000, Lucius & Lucius, Stuttgart.
- Nöldeke, G. and Schmidt, K.M. (1995) "Option contracts and renegotiation: a solution to the hold-up problem," *RAND Journal of Economics* 26, 163-179.

- Packard, D. (1995) *The HP way*, HarperCollins, New York.
- Pittman, T.S., Davey, M.E., Alafat, K.A., Wetherill, K.V., and Kramer, N.A. (1980) "Informational versus controlling verbal rewards," *Personality and Social Psychology Bulletin* 6, 228-233.
- Plant, R. and Ryan, R.M. (1985) "Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: an investigation of internally controlling styles," *Journal of Personality* 53, 435-449.
- Rabin, M. (1993) "Incorporating fairness into game theory and economics," *American Economic Review* 83, 1281-1302.
- Sliwka, D. (2003) "On the hidden costs of incentive schemes," IZA Discussion Paper 844, Bonn.
- Spier, K. (1992) "Incomplete contracts in a model with adverse selection and exogenous costs of enforcement," *RAND Journal of Economics* 23, 432-443.

Appendix 1

Table A1: Agents' choices of x if the choice set is restricted by the principal and if it is not restricted by the principal

Agent	x -choice if restricted	x -choice if not restricted	Type classification	Difference restricted minus not restricted
1	11	0	selfish	11
2	10	0	selfish	10
3	10	0	selfish	10
4	10	0	selfish	10
5	10	0	selfish	10
6	10	0	selfish	10
7	10	0	selfish	10
8	10	0	selfish	10
9	10	0	selfish	10
10	10	0	selfish	10
11	10	0	selfish	10
12	10	0	selfish	10
13	10	0	selfish	10
14	20	10	other	10
15	14	7	other	7
16	15	8	other	7
17	10	5	inequity averse	5
18	10	5	inequity averse	5
19	10	10	inequity averse	0
20	10	10	inequity averse	0
21	10	10	inequity averse	0
22	10	10	inequity averse	0
23	10	10	inequity averse	0
24	12	12	inequity averse	0
25	18	18	inequity averse	0
26	20	20	inequity averse	0
27	35	35	inequity averse	0
28	40	40	inequity averse	0
29	40	40	inequity averse	0
30	40	40	inequity averse	0
31	40	40	inequity averse	0
32	15	17	control averse	-2
33	18	20	control averse	-2
34	18	20	control averse	-2
35	29	31	control averse	-2
36	14	18	control averse	-4
37	10	15	control averse	-5
38	35	40	control averse	-5
39	35	40	control averse	-5
40	35	40	control averse	-5

Table A1 continued

Agent	x -choice if restricted	x -choice if not restricted	Type classification	Difference restricted minus not restricted
41	12	20	control averse	-8
42	10	20	control averse	-10
43	10	20	control averse	-10
44	10	20	control averse	-10
45	10	20	control averse	-10
46	10	20	control averse	-10
47	10	20	control averse	-10
48	20	30	control averse	-10
49	30	40	control averse	-10
50	30	40	control averse	-10
51	30	40	control averse	-10
52	35	45	control averse	-10
53	40	50	control averse	-10
54	95	105	control averse	-10
55	13	24	control averse	-11
56	10	25	control averse	-15
57	10	25	control averse	-15
58	10	25	control averse	-15
59	15	30	control averse	-15
60	12	30	control averse	-18
61	10	30	control averse	-20
62	10	30	control averse	-20
63	10	30	control averse	-20
64	10	30	control averse	-20
65	15	35	control averse	-20
66	20	40	control averse	-20
67	11	40	control averse	-29
68	10	40	control averse	-30
69	10	40	control averse	-30
70	10	40	control averse	-30
71	10	40	control averse	-30
72	10	40	control averse	-30