The Hidden Costs of Control

By ARMIN FALK AND MICHAEL KOSFELD*

We analyze the consequences of control on motivation in an experimental principal-agent game, where the principal can control the agent by implementing a minimum performance requirement before the agent chooses a productive activity. Our results show that control entails hidden costs since most agents reduce their performance as a response to the principal's controlling decision. Overall, the effect of control on the principal's payoff is nonmonotonic. When asked for their emotional perception of control, most agents who react negatively say that they perceive the controlling decision as a signal of distrust and a limitation of their choice autonomy. (JEL D82, Z13)

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. This paper analyzes how the agent perceives the principal's decision to control 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 principal's interest. 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 simple and parsimonious setup. 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 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 various 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. If, however, 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 negative signal of distrust. In addition, these agents might also assume that the principal has low expectations. Our main results, in fact, confirm the hypothesis that control has an adverse effect on agents' performance. We find that a majority of the agents in our experiment choose a *lower x* if the principal restricts rather than trusts them. We vary the level of x in different treatments. While we find evidence for the hidden costs of control in all treatments, the net effect on profits turns out to be nonmonotonic. For low levels of x, control generates significantly lower profits than trust; as x rises, however, control eventually breaks even. In a detailed analysis of agents' motives behind control aversion, we find that agents seem to believe that principals who control expect to receive less than those who don't, and that agents' beliefs correlate positively with their behavior. When asked for their perception of control, most agents indicate that they perceive the decision to control as a signal of distrust and a limitation of their choice autonomy. Given the hidden costs of control, principals earn more if they trust their agents than if they control. Most of the principals in our experiment seem to understand the adverse effect of control. The majority decides not to restrict the agent's choice set but to trust that the agent will perform well voluntarily. Principals who control hold 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" (Niklas Luhmann, 1968).[1]

We conducted several control treatments both to test the robustness of our results and to isolate the precise impact of control on agents' motivation. In one of these treatments, the constraint x is given exogenously, i.e., the agent's choice set is identical to that in the main experiment when the principal decides to control. Since it is not the principal who implements x, however, distrust and control are not at issue in this treatment. Assuming that the principal's controlling decision is behaviorally relevant to the agent, if the principal decides to control in the main treatment, x should be lower than that in the control treatment. This is what we find. In a further treatment, we embed the principal's control decision in a standard gift-exchange game. A principal in this treatment not only decides whether to control his agent but also determines his agent's wage. In line with previous studies, we find that agents' x-choices are generally increasing in the wage payment, i.e., agents act reciprocally to the principal's wage choice. Yet, if the principal controls, the agent's reciprocal inclination is lower than if the principal does not control. The result shows that even if the principal can use additional instruments to motivate the agent, control continues to have a negative impact on agents' performance. To further test the robustness of our findings, we conducted a questionnaire where subjects were asked to state their work motivation in various everyday workplace scenarios. Similar to our lab evidence, we find that self-reported work motivation varies significantly with the extent to which agents are exposed to control.

Taken together, our results show that the use of control and 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 help. Similarly, characteristics of the workplace environment that limit freedom of choice and signal low expectations, such as high levels of monitoring and surveillance, may be equally counterproductive.

In fact, the hidden costs of control offer a

---

[1] The 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 store-

rooms were a symbol of trust, a trust that is central to the way HP does business" (David Packard, 1995, p. 135).

psychological rationale for the 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 that 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 incomplete *deliberately*. 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 (Philippe Aghion et al., 1994; Georg Nöldeke and Klaus M. Schmidt, 1995). In other cases, incomplete contracts may actually be superior to more complete contracts. Franklin Allen and Douglas Gale (1992) and Kathryn E. Spier (1992), for example, argue that offering a more complete contract may lead the other party to draw negative inferences about the first party's type. B. Douglas Bernheim and Michael D. Whinston (1998) show in a repeated game setting that if some obligations are noncontractible, 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 nonopportunistic behavior.

Our paper also contributes to the recent literature dealing with the interaction of psychological and economic incentives (e.g., Bruno S. Frey, 1997; Uri Gneezy and Aldo Rustichini, 2000a, b; Frey and Reto 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 agent's choice set 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 (Ernst Fehr and Simon Gächter, 2002; Fehr and Bettina Rockenbach, 2003; Fehr and John A. List, 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 pays in our experiment only because some agents are intrinsically trustworthy and react negatively to the implementation of control. A further distinction of our paper is that we are able to obtain individual-level information on the agent's behavior rather than aggregate results alone. In particular, we can analyze agents' heterogeneity with regard to their behavioral motivation by distinguishing between agents who react positively, neutrally, and negatively to control. Finally, our study highlights previously untested mechanisms for possible dysfunctional effects of incentives. Earlier 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; Roland Bénabou and Jean 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. Two recent theoretical papers that investigate the optimal monitoring of heterogenous agents are Dirk Sliwka (2003) and Andrea Ichino and Gerd Mühlheusser (2004).[2]

---

[2] Harry G. 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 work-

The paper is organized as follows. Details of the experimental design are explained in the following section. Section II discusses the behavioral predictions. Our main results are presented in Section III. Section IV reports data from the gift-exchange experiment. Finally, in Section V we present the results from a questionnaire illustrating the applications and economic importance of the hidden costs of control for real-life labor relations. Section VI concludes.

I. Experimental Design

A. *Main Treatments*

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 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 $x \in \{\underline{x}, \underline{x} + 1, \dots, 120\}$, or he can leave the choice set unrestricted to $x \in \{0, 1, \dots, 120\}$. The parameter $\underline{x}$ varies across treatments (see below). Thus, the principal can control the agent's decision environment, thereby guaranteeing a minimal payoff of $2\underline{x}$, or he can leave

the decision completely up to the agent, trusting that the latter will not choose x below $\underline{x}$.

We conjecture that the impact of control depends on the level of control, which is measured by the parameter $\underline{x}$ in our game. If $\underline{x}$ is low, for example, the positive effect of control is likely to be marginal; the hidden costs of control on the other hand (if they exist) may be substantial. As $\underline{x}$ increases, by definition the positive effect of control rises, since agents cannot choose below $\underline{x}$. In consequence, it may turn out that the overall net effect of control is actually nonmonotonic. For low levels of $\underline{x}$, control might generate a lower payoff than no-control; as $\underline{x}$ rises, the payoff from control increases; eventually, control breaks even.[3] To see whether this hypothesis is correct, we implemented a low, a medium, and a high control treatment. In the low control treatment (C5), the principal can only force the agent to choose $x \geq 5$. He is able to enforce twice as much, i.e., $\underline{x} = 10$, in the medium control treatment (C10). Finally, he can enforce a minimum of $\underline{x} = 20$ in the high control treatment (C20). We made use of the strategy method to elicit the agents' choices in the experiment in each of these treatments. Before knowing their principal's actual decision, the agents had to decide on x for both possible cases. On a computer screen, they were asked to choose $x \in \{\underline{x}, \underline{x} + 1, \dots, 120\}$ in case their principal forces them to choose at least $\underline{x}$, 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.

B. *Control Treatments*

We also ran several control treatments to check for the robustness, for possible interpretations, as well as for the validity of our results in a more general economic setting. The medium control treatment C10 serves as the main basis of comparison for all these treatments.

Our first control treatment (SR10) tests whether agents' behavior depends on the strategy method as an elicitation procedure. If the experience of being controlled is emotionally

---

ing hours. The study uses data from 116 executives of medium-sized Dutch firms. While the author finds that higher monitoring actually correlates with fewer working hours in some cases, the causal relationship between the two variables is unclear. A related literature in social psychology comes from Thane S. Pittman et al. (1980), Mark R. Lepper and David Green (1975), Robert Plant and Richard M. Ryan (1985), and Michael E. Enzle and Sharon C. Anderson (1993). None of these papers studies control in a principal-agent relation, however.

[3] Gneezy (2004) provides recent evidence for the nonmonotonicity of incentives in a proposer-responder game.

important, it may well be that the aversion against control is even stronger if agents decide after having learned whether they are controlled or not. On the other hand, it may be that the strategy method places too much emphasis on the control-versus-trust decision, thereby inducing agents to overstate their dislike for control. To rule out the possibility that our results are an artifact of the strategy method, we therefore applied the specific response method in treatment SR10. Principals first set a minimum of zero or $x = 10$; then agents learn their principal's choice before making their decision. In consequence, each agent makes only one decision in this treatment—either under control or under no control.

To test whether it is really the principal's control decision that affects agents' motivation, we implemented a second control treatment where the constraint $x \geq 10$ is given exogenously to the agent (EX10). In treatment EX10, the principal and the agent play only the subgame of the game in treatment C10 that follows the restriction choice of the principal. As before, the agent is endowed with 120 points. He chooses a productive activity $x \in \{10, 11, ... , 120\}$. Payoffs are $120 - x$ for the agent and $2x$ for the principal. The principal makes no decision.

Finally, we explore the validity of our results in a more general economic setting in a third control treatment. The main advantage of our simple principal-agent game is that it allows us to study the hidden costs of control in a setting uncluttered with other factors. The potential disadvantage is that it may exclude essential features of employment relations that might be relevant for the robustness of our results. In practice, principals clearly do more than determine the agent's choice set. Most importantly, they also set the agent's wage. To check whether the introduction of wages has an impact on the hidden costs of control, we implemented a gift-exchange treatment (GE10) where the principal not only determines the agent's minimum level of x, but also pays the agent a wage. More precisely, in the first stage of treatment GE10, the principal chooses a wage $w \in \{10, 30, 60, 120\}$ and decides whether to force the agent to choose $x \geq 10$. Depending on the principal's decision, the agent in the second stage chooses $x \in \{10, 11, ... , 120\}$ if the principal imposes control and $x \in \{0, 1, ... ,$

120$\}$ if the principal does not impose control in the first stage. Payoffs are $\pi_p = 2x - w$ for the principal and $\pi_a = w - x$ for the agent. Note that the game in treatment GE10 is a straightforward modification of our main principal-agent game (C10), the only difference being that players' payoffs are now also determined by the principal's wage choice. We made use of the strategy method to elicit agents' choices in this treatment; i.e., for each possible wage w agents made a choice of x under both the condition where the principal does and does not implement a minimum of 10.

C. Procedural Details

Irrespective of the treatment, each principal-agent game was played one-shot. All treatments were framed in a neutral manner.[4] All experiments were computerized using the software "z-Tree" (Urs 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. A total of 804 subjects participated in the experiment, half of them as agents, the other half as principals. No subject participated in more than one treatment. A total of 140, 144, and 134 subjects participated in the main treatments C5, C10, and C20, respectively; 246 subjects participated in treatment SR10. Finally, 72 subjects participated in treatment EX10 and 68 participated in treatment GE10. Each session lasted, on average, between 40 and 50 minutes. On average, a subject earned CHF 25.11 (about $20) in the experiment.

D. Questionnaire Study

In addition to our experiment, we also conducted a questionnaire to illustrate the variety of applications and economic importance of the hidden costs of control in typical real-life labor relations. We used vignettes in the questionnaire where subjects were exposed to everyday workplace scenarios. This allows changing conditions along the control-trust dimension, hold-

---

[4] In general, subjects were called "participant A" and "participant B." In the gift-exchange treatment participants were called "employee" and "employer." We avoided value-laden terms like "trust" or "distrust."

ing the general workplace scenario constant. A total of 403 subjects participated in the questionnaire study. None of these subjects participated in any treatment of the experiment. We provide more information about the questionnaire in Section V.

II. Behavioral Predictions

The behavioral predictions for our experiments depend on the assumptions 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 $x = \underline{x}$ if they are restricted and $x = 0$ otherwise. In consequence, principals who want to maximize their payoffs 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 Colin F. Camerer, 2003). In the dictator game, for example, proposers often give positive amounts of money (Robert Forsythe et al., 1994; Alvin E. Roth, 1995). These findings were recently explained in terms of subjects' concern for equity and efficiency (Fehr and Schmidt, 1999; Gary E. Bolton and Axel Ockenfels, 2000; Gary Charness and Matthew Rabin, 2002). 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. If, however, the agent's choice exceeds $\underline{x}$ (i.e., the agent has a relatively strong preference for equitable outcomes), social-preference theories imply that his choice is in fact independent of whether a principal controls him or not. This follows because, in these models, social preferences are based on payoff distributions and the constraint $x \geq \underline{x}$ is not binding. Only agents with a weak social preference, who choose $0 < x < \underline{x}$ if not controlled, will choose a higher performance (namely, $x = \underline{x}$) 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 agents with social preferences, and hence should always restrict the agent's choice set.[5]

Now suppose that agents are intrinsically motivated to perform in the principal's interest but reduce their performance in response to being controlled. A potential reason why we might see this behavioral reaction is that controlling implicitly signals that the principal does not expect the agent to perform well. Some agents might perceive this as a signal of distrust in their intrinsic motivation; others might take it as an indication of how much voluntary performance the principal expects. In both cases, the agent reacts by choosing a lower performance than he would have chosen if the principal had not controlled the agent. If faced with this type of behavioral motivation, the optimal strategy for the principal is thus the opposite of that described above: a principal is better off trusting and not controlling the agent. Of course, if the enforceable level of control $\underline{x}$ is higher than what the agent is willing to choose voluntarily, controlling also pays if the agent is control averse. Therefore, the payoff of controlling is likely to be nonmonotonic. If $\underline{x}$ is relatively low, trust pays; as $\underline{x}$ increases, control may become the better strategy.

With regard to the exogenous-control treatment EX10, note that both selfishness and social preferences predict no difference in agents' behavior between treatment EX10 and the subgame in treatment C10 that follows the restriction decision. In contrast, if agents are averse to control, x-choices should be higher in

---

[5] Formally, according to the model of Fehr and Schmidt (1999), for example, 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 nonlinear versions of the model. Other existing theories of social preferences are not only based on payoff distributions but also take players' intentions into account (Rabin, 1993; Martin Dufwenberg and Georg Kirchsteiger, 2004; Falk and Fischbacher, 2006). None of these theories, however, predicts that in equilibrium the principal refrains from controlling the agent's choice set, and the agent chooses a lower x if he is controlled than if he is not controlled by the principal. The main reason is that if the principal's decision not to control the agent is perceived as a kind action (which must be the case if the agent's response is positive), controlling must be even kinder, because the agent's utility is higher as his performance is lower.

treatment EX10 than in the corresponding subgame of treatment C10. The reason is that while the strategy set and the payoff consequences are identical, the principal does not actively choose the constraint $\underline{x}$; instead, it is imposed exogenously in treatment EX10. Hence, the constraint cannot signal anything.

Finally, the principal also sets the agent's wage in the gift-exchange treatment GE10. Based on results from previous studies (Fehr et al., 1993; Fehr et al., 1997; Fehr and Falk, 1999; Gneezy, 2003), we expect reciprocity considerations to become relevant in this treatment. That is, agents' motivation to act in the principals' interest should increase with the wage payment. This hypothesis is also in line with social-preference theories. As is true for our main treatments, however, social-preference theories predict that control is optimal in the GE10 treatment because the controlling decision does not affect agents' reciprocity. In contrast, if agents are averse to control, control and reciprocity may interact in a nontrivial way. If wages are low, even reciprocal agents will not choose a high x. Therefore it pays for the principal to control. For higher wages, however, control may no longer be optimal. The reason is that by setting a high wage and waiving control, the principal unambiguously signals his kindness and his trust in the agent's performance. The implementation of control in this case, on the other hand, signals less trust and kindness. In consequence, the agent's reciprocal inclination may be less pronounced if the principal controls rather than trusts. This may result in lower payoffs.

III. Main Results

In this section, we present the results from our main treatments. We first report agents' behavior on the aggregate and on the individual level. We also show the robustness of our results with regard to the elicitation procedure. We then turn to the principals. Finally, we discuss possible interpretations of our findings and explore agents' behavioral motives.

A. *Agents' Behavior*

Our first result concerns the hidden costs of control.

RESULT 1: *We observe hidden costs of control in all main treatments (C5, C10, and C20).*

Support for Result 1 comes from Figure 1A–C. The figure shows how agents' choices are cumulatively distributed given the decision of the principal in the three treatments. If there were no hidden costs of control, the cumulative distributions of agents' choices, in case the principal controls (black points), would coincide with the cumulative distributions if the principal does not control (white points) for all $x \geq \underline{x}$. Obviously, this is not the case in any of the three treatments. For each value of $x \geq \underline{x}$, there are always strictly more agents in the no-control condition who choose at least that value of x than in the control condition. For example, 67 percent of the agents in treatment C5 (Figure 1A) choose $x > 10$ when they are free to choose voluntarily, and only 20 percent choose $x \leq 5$. In sharp contrast, only 30 percent choose $x > 10$ if they are forced to choose at least 5, and 51 percent choose exactly the minimum $x = 5$; 34 percent of the agents in this treatment choose the payoff equalizing level of $x = 40$ if they are not controlled by the principal, yet only 9 percent do so if they are controlled. A similar picture emerges in treatments C10 and C20 shown in Figure 1B and 1C. Independent of the level of $\underline{x}$, we see relatively high x-choices from the agents if they are in an uncontrolled environment, whereas the mass of x-choices is centered at $\underline{x}$ if the principal restricts the agents' choice set.

The shift in the distributions is highly significant in all three treatments. To test this, we modify the distribution of x-choices in the no-control case such that any observation $x < \underline{x}$ is set equal to $\underline{x}$. Under the assumption that control has no behavioral impact, the resulting modified distributions should be the same as the ones generated by control. This can be rejected for all treatments (Wilcoxon signed rank test, $p < 0.001$).[6]

While control entails hidden costs, by definition it may also generate benefits. Forcing agents to choose at least $\underline{x}$ implies that any choice below $\underline{x}$ is ruled out for sure. Our next result shows how the costs and benefits translate into profits.

---

[6] We report the results of two-sided tests throughout the paper.
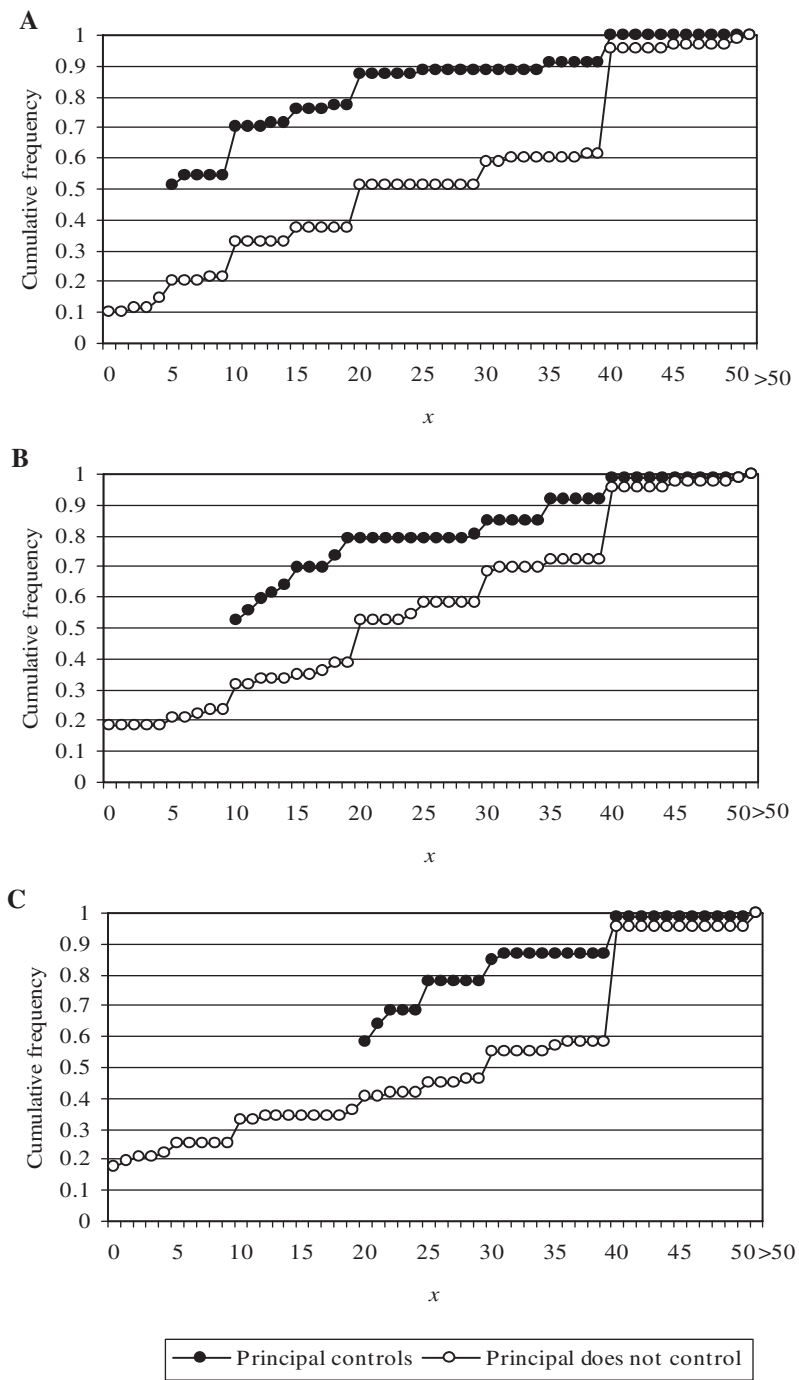
FIGURE 1. CUMULATIVE DISTRIBUTION OF AGENTS' CHOICES IN TREATMENT C5 (PANEL A), C10 (PANEL B), AND C20 (PANEL C)

Notes: The figure shows all observations $x \leq 50$. In each treatment, there were a few x-choices above 50. These observations are summarized as $x > 50$.

TABLE 1—AGENTS' CHOICES DEPENDENT ON THE
PRINCIPAL'S DECISION

		Treatment		
		C5	C10	C20
Principal controls	Average	12.2	17.5	25.4
	Median	5	10	20
Principal does not control	Average	25.1	23.0	26.7
	Median	20	20	30

Note: Number of observations: $n = 70(C5)$, $n = 72(C10)$, $n = 67(C20)$.

RESULT 2: *The hidden costs of control outweigh the benefits in all main treatments. Average performance is higher if the principal does not control than if he does so. These differences are significant in the C5 and the C10 treatment but not in the C20 treatment.*

Result 2 is supported by Table 1. Table 1 shows agents' average and median performance in our main treatments if the principal does and does not restrict the agent's choice set. This table conveys a clear message. In all treatments, both average and median performance levels are higher in the no-control case than in the control case. Depending on the level of x, the differences are actually quite large. In treatment C5, agents make choices that are on average more than twice as high if they are not controlled by their principal than if they are controlled; in treatment C10 average choices are 31 percent higher. Median differences are even stronger. As we have seen in Figure 1, if the principal sets a minimum of x, more than half of the agents choose exactly the minimum; hence median performance equals 5, 10, and 20 in treatments C5, C10, and C20 if the principal controls. The median performance in case the principal trusts, on the other hand, is 20, 20, and 30. These performance differences are significant in the C5 and the C10 treatment (Wilcoxon signed rank test, $p < 0.001$). Agents' choices in the C20 treatment are also slightly higher for principals who trust, but not significantly so (Wilcoxon signed rank test, $p = 0.957$).

If we compare agents' average performance under control and no-control across the treatments (i.e., for different levels of x), the data reveal a further interesting finding. While the enforceable level of control x plays a crucial role if control is actually implemented, it plays no important role if the principal decides to trust. Agents do not behave differently if the principal refrains from implementing a minimum of $x = 5$, compared to if he refrains from implementing a minimum of 10, or 20 (Kruskal-Wallis test, $p = 0.550$). This suggests that agents seem to punish the principal's decision to control rather than reward his decision to trust. We will return to this issue in Section IIIC, when we compare agents' behavior to the data from the exogenous-control treatment EX10.

Result 1 emphasizes the hidden costs of control caused by the existence of intrinsically motivated agents choosing a lower performance if controlled than otherwise. Result 2 shows that control also has some benefits, since selfish agents are forced to choose a higher level of x if they are controlled than if they are not controlled. In the following, we provide more information about agents' heterogeneity with regard to their behavioral reaction to control. Since we used the strategy method, we are able to quantify how many agents reacted negatively, positively, or in a neutral way, depending on the principal's decision.

RESULT 3: *There is a strong heterogeneity among the agents in all main treatments. We see agents who react positively, neutrally, or negatively to the principal's implementation of control. The last group is always the majority.*

Support for Result 3 comes from Table 2. This table shows the absolute and relative frequencies of agents who react positively, neutrally, or negatively to the implementation of control in the three main treatments, together with their average choices given the principal's decision. If all agents were selfishly motivated, reactions would be positive, and the average performance would equal x if controlled and zero otherwise. Table 2 instead reveals that, depending on the treatment, only between 20 and 37 percent of the agents show a positive performance reaction to control. On average, these agents choose slightly higher levels of x than the required minimum. For the second group of agents, the decision of the principal is irrelevant in their choice of x. This group consists of between 16 and 21 percent of the agents, depending on the treatment. Finally, the largest group in each of the three treatments consists of individuals who show a negative response to

TABLE 2—HETEROGENEITY OF AGENTS' BEHAVIORAL REACTION TO CONTROL

| | Treatment | | | | | | | | |
| | C5 | | | C10 | | | C20 | | |
	Positive	Neutral	Negative	Positive	Neutral	Negative	Positive	Neutral	Negative
Number of agents	14	11	45	18	13	41	25	14	28
Relative share	0.20	0.16	0.64	0.25	0.18	0.57	0.37	0.21	0.42
Average x if controlled	10.2	22.3	10.3	11.1	22.7	18.7	21.9	39.4	21.5
Average x if not controlled	4.8	22.3	32.1	1.9	22.7	32.3	4.9	39.4	39.8

being restricted in their choice set; 64, 57, and 42 percent of the agents reveal this behavioral pattern in treatments C5, C10, and C20, respectively. They are responsible for why, on average, it does not pay for the principal to control. Their average choice of x if controlled is 10.3, 18.7, and 21.5 for $\underline{x}$ equal to 5, 10, and 20, respectively. If they can decide freely, on the other hand, their average choice is 32.1, 32.3, and 39.8.

In sum, our results show that the costs of control are substantial. The results do *not* suggest, however, that trust is always better than control. In fact, they show that the hidden costs and benefits of control depend on various factors. First, they depend on the relative frequency of agents' types. When the number of opportunistic agents with a low intrinsic motivation to perform is relatively high, controlling only generates minor costs and trusting is likely to be suboptimal. Second, the level of the explicit incentives is important. To illustrate, compare the principals' average profits in the C5, C10, and C20 treatments. In the C5 treatment, controlling lowers the motivation of many agents, while the benefits of control are relatively small given the little enforcement power of $\underline{x} = 5$. As a consequence, the difference in average profits between controlling and trusting amounts to $24.4 - 50.2 = -25.8$. Controlling also lowers motivation in the C10 treatment, but the enforcement power is stronger. As a result, the net benefit of control is now $35.0 - 46.0 = -11.0$. Finally, in the C20 treatment there are again negative effects of control, but due to the much stronger enforcement power, average profits differ insignificantly $(50.8 - 53.4 = -2.6)$. We suspect that control will outperform trust for even higher levels of $\underline{x}$. In this sense, our results suggest that the relation between the principal's

payoff and the strength of incentives is indeed nonmonotonic: if the principal has only weak incentives at his disposal, it may be better to trust, since controlling lowers motivation of the intrinsically motivated agents but increases the performance of opportunistic agents only marginally. As controlling technologies and enforcement power increase in strength, however, the disciplining effect eventually dominates the negative effect on motivation.

Before we turn to the principals, we test the robustness of our results with regard to the elicitation of agents' behavior. We elicited agents' choices with the help of the strategy method in treatments C5, C10, and C20. This was predominantly done to enable us to study potentially different types of agents (Result 3). To rule out the possibility that agents' behavior is an artifact of the strategy method, we conducted a control treatment (SR10), where we used the specific response method instead. In this treatment, agents decided only after having learned of the principal's decision. Our results do not indicate any effect of using the strategy method versus the specific response method in our context. The average (median) choice of an agent is 23.6 (20) if the principal does not implement a minimum level of 10 in the specific response treatment. Using the strategy method, we get an average (median) of 23.0 (20). These differences are not significant (Mann-Whitney test, $p = 0.822$). If the principal decides to set a minimum of 10, the average (median) is 19.6 (10.5) if we use the specific response method, and 17.5 (10) in the strategy method. Again, there is no significant difference (Mann-Whitney test, $p = 0.589$). Finally, the hidden costs of control, i.e., the difference in the distributions between the control and the no-control con-

TABLE 3—PRINCIPALS' BEHAVIOR AND BELIEFS

| | Treatment | | | | | |
| | C5 | | C10 | | C20 | |
	Control	Trust	Control	Trust	Control	Trust
Relative share	0.26	0.74	0.29	0.71	0.48	0.52
Average belief of x	17.8	29.6	19.4	25.7	25.3	34.1
Average counterfactual belief of x	12.8	14.9	—	—	10.3	23.0
Average x actually chosen	12.2	25.1	17.5	23.0	25.4	26.7
Are beliefs "correct"?	Yes	Yes	Yes	Yes	Yes	No

Notes: Counterfactual beliefs were elicited only in treatments C5 and C20. Beliefs are "correct" if the Mann-Whitney test does not reject the hypothesis that actual choices and corresponding beliefs are the same ($p > 0.1$).

dition, are significant in the SR10 treatment as well (Mann-Whitney test, $p = 0.007$).[7]

B. *Principals' Behavior*

We now turn to the principals. Given the average responses of the agents, it is clearly not optimal for payoff-maximizing principals to control their agent, in particular in treatments C5 and C10. Our next result shows that principals seem to have understood this.

RESULT 4: *The majority of the principals chooses not to control the agent.*

Support for Result 4 comes from Table 3, which reports the principals' decisions. In all treatments, the majority of principals decides to trust rather than to control. The differences are largest in the C5 treatment, where only 26 percent of the principals control and 74 percent trust. The respective numbers are 29 and 71 percent in the C10 treatment. These differences are statistically significant for both treatments at any conventional level (Binomial test, $p < 0.001$ in each treatment). The data suggest that principals anticipate possible adverse effects of controlling their agents because they voluntarily give up the option to guarantee $2x$ points for themselves. The differences are smallest and insignificant in the C20 treatment, where roughly half of the principals controlled while the other half did not. In light of the performance levels actually chosen, these results

make sense: the larger the payoff differences between trusting and controlling, the larger is the fraction of principals who trust.

The results shown in Table 3 suggest that even though a majority of principals chose an optimal strategy, some did not. This raises the question why some of the principals chose to control while others trusted. To better understand the principals' decisions, we study principals' beliefs. These beliefs were elicited right after principals had made their decision. In all treatments, we asked them to state their expectation about the agent's choice of x.[8] In the C5 and the C20 treatments, we also asked them about their counterfactual beliefs, i.e., what those principals who controlled think they would have gotten had they not controlled, and vice versa. The main finding is summarized in our next result.

RESULT 5: *Principals who control have lower expectations about x than principals who do not. Expectations coincide with agents' average performance in most of the cases.*

Principals' beliefs are displayed in Table 3. Principals who control expect to get less than those who trust in all main treatments. In the C5 treatment, e.g., principals expect to receive on average 17.8 if they control and 29.6 if they trust. Similarly, principals who control expect 19.4 and 25.3 on average in treatments C10 and C20, whereas their trusting counterparts in the same situation expect on average 25.7 and 34.1. The differences in beliefs are statistically significant in each of the treatments (Mann-Whitney

---

[7] As above, we account for the difference in the support of x by setting each observation $x < 10$ in the no-control case equal to 10.

[8] Beliefs were not paid.

test, $p < 0.007$). The result suggests that principals share different views about their agents' cooperativeness. Those who trust are rather optimistic, while those who control seem to be rather pessimistic. It is necessary to consider also principals' counterfactual beliefs, however, in order to substantiate this claim. Do those principals who control expect to receive less if they trust, and vice versa? The answer is affirmative. In both the C5 and the C20 treatments, principals who think that they get more if they control than if they trust choose to control, and vice versa. This implies two things. First, principals act rationally conditional on their beliefs. Second, there are in fact two types of principals. Those who are optimistic believe that trust pays, while those who are pessimistic expect control to be the better strategy.[9]

A comparison of the principals' expectations with agents' actual choices reveals that expectations are indeed quite accurate. For example, principals who control in treatment C10 expect to get 19.4 on average when they in fact earn 17.5. At the same time, principals who trust in this treatment expect on average 25.7, and indeed they receive 23.0. Table 3 shows that principals' expectations are in general a bit higher than agents' actual performance. With the exception of the principals who trust in treatment C20, however, a Mann-Whitney test does not reject the hypothesis that beliefs and actions are the same both if principals do not control and if they do so ($p > 0.13$ in each of these cases). In the C20 treatment, principals who trust expect to receive 34.1 but in fact they get only 26.7 (Mann-Whitney test, $p = 0.071$). This optimism might have been due to the expectation that not choosing to control if one could have enforced a relatively high amount will be rewarded by particularly high choices of x. Apparently, this is not the case. We will return to this issue in the following section.

In our experiment, the principals' beliefs and

the agents' actual performance 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 their agents' relatively low average performance. 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 rely on control. These managers will not be surprised to see performance at the minimum, confirming their pessimistic beliefs. The locked storeroom policy of General Electric in the 1930s 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 nonminimal performance, substantiating the initial beliefs held by the managers.

C. *Exploring Agents' Motives: What Drives the Hidden Costs of Control?*

In the preceding section we showed that control has an adverse effect on many agents' performance. We now explore possible interpretations of this finding. We first present results of the EX10 treatment that allows checking whether the control decision of the principal really drives our results in the main treatments. Second, we examine whether agents' beliefs about the principal's expectations play a role in agents' behavior. Finally, we report agents' answers to a free-form questionnaire, in which they were asked to describe their emotional perception of being controlled by the principal.

Strategy Space versus Control Decision.—In principle, the negative effect of control in our main treatments could be due to the mere difference in the agents' strategy space, i.e., to the difference in the support of x between the control and the no-control condition. To isolate the effect of the principal's control decision, we compare agents' choices in the subgame that follows the control decision in treatment C10

---

[9] Interestingly, both types of principals agree in their expectations about agents' behavior in case of control. If we compare the beliefs of principals who control with the counterfactual beliefs of principals who trust (treatments C5 and C20), a Mann-Whitney test does not reject the hypothesis that beliefs are the same ($p > 0.4$ in both treatments). The actual difference between principals arises when it comes to their expectation about agents' performance in case they are trusted.

with agents' choices in treatment EX10 where subjects play only this subgame. Agents' strategy space is identical (10 to 120) in both situations. In the subgame of treatment C10, however, the principal decided to control the agent; in treatment EX10, x is imposed exogenously, i.e., any control decision of the principal is absent. If the controlling decision of the principal generates the negative effect, the x-choices in the EX10 treatment should exceed those in the C10 subgame following the control decision. This is what we find. If agents are exogenously constrained to choose $x \geq \underline{x} = 10$, the average (median) choice of an agent is 28.7 (20) compared to 17.5 (10) if the principal implements the same restriction endogenously. These differences are statistically significant (Mann-Whitney test, $p < 0.001$). At the same time, there is no significant difference between agents' choices in the EX10 treatment and agents' choices in the subgame of the C10 treatment when the principal trusts, i.e., does not impose control (Mann-Whitney test, $p = 0.523$).[10] Thus, agents' behavior is indeed a reaction to the principal's decision to control.[11] Moreover, since the average performance in the trust subgame of the C10 treatment is not higher than in the EX10 treatment, it seems that agents punish the principal's controlling decision rather than rewarding his decision to trust. This is also in line with our finding from the main treatments showing that agents do not respond differently to the waiving of different levels of control $\underline{x}$ (see Table 1).

Agents' Beliefs about Principals' Expectations.—The fact that agents react to the principal's decision suggests that the latter may provide a signal—for example, a signal about his expectations concerning the agent's performance. Recall that principals who control have lower expectations than principals who trust (Result 5). If agents understand this, it seems plausible that agents differ in their belief about the principal's expectations: agents who are controlled probably believe that the principal has lower expectations with regard to x compared to agents who are free to choose voluntarily. Moreover, agents' beliefs are likely to affect their behavior. Charness and Dufwenberg (2004), for example, provide evidence that guilt aversion is an important motive in principal-agent relations. Agents who believe that the principal has low expectations may feel less "guilty" when choosing a low x than agents who believe that the principal has high expectations. In consequence, low expectations generate low performance. At the same time, the agent can also interpret low expectations in the sense that the principal distrusts him to perform well. If the agent doesn't like to be distrusted, because he regards himself as a trustworthy person, the principal's low expectations similarly reduce the agent's performance. In order to check whether agents' beliefs play a role, we asked agents the following question in the control treatment SR10, where we used the specific response method: "What do you think were the expectations of participant B (principal) concerning your transfer decision?"[12] The question was asked directly after the agent had made his decision. The results are as follows. The average (median) belief of an agent who is forced to choose $x \geq 10$ is 26.1 (30) compared to 35.9 (40) if the agent is free to choose voluntarily (Mann-Whitney test, $p = 0.075$). Thus, agents who are controlled think that the principal has lower expectations than do agents who are not controlled. To test whether beliefs also affect agents' behavior, we ran a linear regression of agents' transfers on their beliefs, controlling for the principal's decision. The regression shows that agents' transfers and beliefs are significantly and positively correlated (coef. = 0.2962, $p = 0.050$). Thus, the less an agent believes that his principal expects, the less he is willing to perform in the principals' interest.

Agents' Emotional Perception of Control.—To shed further light on agents' perception of being controlled by the principal, we asked all agents in our main treatments—C5, C10, and C20—

---

[10] Again, we account for the difference in the support of x by setting each observation $x < 10$ in the no-control case equal to 10.

[11] This result can be interpreted in terms of intention. In a treatment where the outcome of a "decision" cannot be attributed to the actor's intention, reciprocal responses are typically weaker (Sally Blount, 1995; Falk et al., 2000, 2003; Charness, 2004).

[12] We asked this question only to the 52 agents who participated in the last three sessions of treatment SR10. Beliefs were not paid.
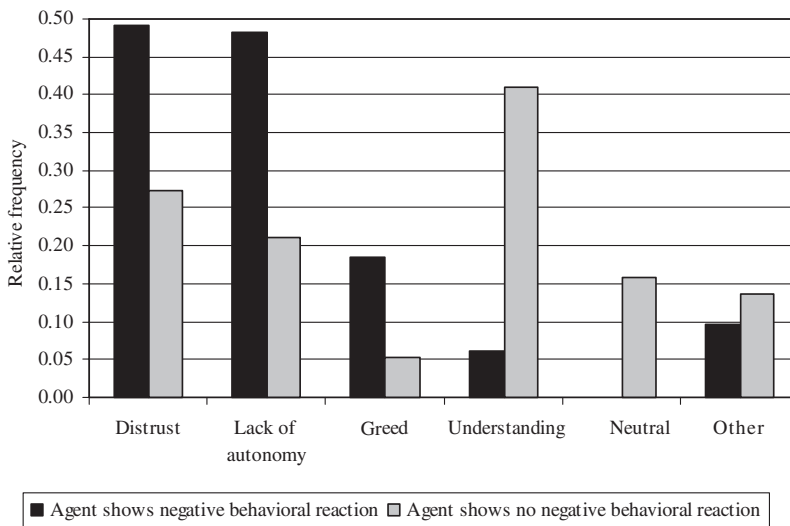
FIGURE 2. AGENTS' EMOTIONAL PERCEPTION OF CONTROL

Notes: Response to: "What do you feel if participant B (principal) forces you to transfer at least x points?" (C5, C10, and C20; data pooled over treatments; $n = 209$).

the following question: "What do you feel if participant B (principal) forces you to transfer at least x points?" The question was asked after all decisions in the experiment were made. Answers were given in free form. It turns out that the answers can be organized in six categories: distrust, lack of autonomy, greed, understanding, neutral, and other.[13] The fact that none of the subjects mentions that he or she feels less guilty about transferring a low x to the principal was quite surprising to us. One possible reason might be that the question was unfavorable to check for guilt aversion. Had we asked a question like: "What would you have felt if you had

transferred the lowest possible amount?," we speculate that we might have found guilt aversion as an additional category. Figure 2 shows the distribution of answers to our question across the six categories for all subjects who participated in the role of an agent in treatments C5, C10, or C20 (data pooled over treatments, $n = 209$). The dark bars show the answers of agents who reacted negatively to control, i.e., who chose a lower x when they were controlled than when they were not controlled by the principal ($n = 114$). The light bars show the answers of the remaining agents who either reacted negatively or who made the same choice of x in both conditions ($n = 95$).

Let us first look at the agents who reacted negatively to control in the experiment. Figure 2 shows that the emotional perception with the highest frequency is distrust; 49 percent of the agents who are willing to transfer a positive amount say that they feel distrusted if the principal forces them to transfer at least x, and that they are hurt by this distrust in their voluntary motivation.[14] Next, a comparable number of agents (48 percent) say that they feel a lack of

---

[13] An answer was categorized as indicating "distrust" if the subject explicitly mentioned a feeling of not being trusted to transfer a positive amount. Category "lack of autonomy" contains all answers, where the subject expressed a negative feeling of being restricted in his or her freedom of choice. "Greed" represents answers where the subject felt that the principal was a greedy or a petty person. Categories "understanding" and "neutral" contain answers where the subject said that he or she understands the principal's decision or where the subject did not express any particular feeling, respectively. Finally, the category "other" contains all answers that could not be classified into one of the other categories. Note that answers may well fall into more than one category. For example, a subject may mention both a feeling of distrust and of a lack of autonomy. In consequence, frequencies generally do not add up to one.

[14] A typical response in this category reads, as one subject puts it: "The fact that he does not trust me to transfer enough points offends my pride."

autonomy and a reduced opportunity for determining the outcome at their own will,[15] and 18 percent feel that the principal seems to be a greedy or a petty person. Finally, only a minority (6 percent) expresses understanding for the principal's decision to restrict their choice set, and none of the agents says that he or she feels neutral.

In contrast, agents who do not react negatively in the experiment mostly express their understanding for the principal's decision (41 percent); 16 percent do not express any particular feeling. Interestingly, many of these agents also perceive the controlling decision of the principal as a signal of distrust (27 percent) or as a lack of autonomy (21 percent). Unlike the previous group, however, these agents do not lower their performance. Either they choose the same value of x independent of the decision of the principal, or they chose a higher x, mainly because the controlling decision forces them to do so. Apparently, the principal's distrust in the agent's performance seems well justified in the latter case.

In sum, our results indicate that the controlling decision really matters. Agents seem to believe that principals who control expect to get less than those who don't (which is correct, as Result 5 shows), and their beliefs positively correlate with their behavior. When asked for their emotional perception of control, most agents who react negatively say that they perceive the controlling decision as a signal of distrust or a limitation of their choice autonomy.[16]

IV. Hidden Costs of Control in the Gift-Exchange Game

The treatments we have discussed thus far are extremely simple, which has the advantage of

allowing us to study hidden costs of control in a setting uncluttered with confounding factors. On the other hand, our setup abstracts from many essential features of employment relations that are potentially relevant and that may interact with hidden costs of control. Most importantly, principals set wages in practice. Since the payment of wages signals something about the principal's expectation about the agent's trustworthiness and willingness to work, wages complicate the perception of control in a nontrivial way. Therefore, in this section, we report the results from our gift-exchange treatment GE10. Recall from Section I that the principal both chooses a wage $w \in \{10, 30, 60, 120\}$ and decides whether to set a minimum $x \geq 10$ in the first stage of this treatment. Depending on the principal's decision, the agent in the second stage chooses $x \in \{10, 11, ... , 120\}$ if the principal imposes control and $x \in \{0, 1, ... , 120\}$ if the principal does not impose control. Payoffs are $\pi_p = 2x - w$ for the principal and $\pi_a = w - x$ for the agent.

The gift-exchange treatment GE10 is essentially the same as our main treatment C10 but allows principals to set wages. We therefore expect reciprocity considerations on the side of the agents to become relevant. Numerous gift-exchange experiments similar to ours have shown that agents provide efforts above the contractually enforceable level, and that efforts are, on average, increasing in the wage payment. The reason is that high wages are perceived as kind or fair, and reciprocal workers respond to this perceived kindness by providing relatively high efforts (Fehr et al., 1993; Fehr et al., 1997; Fehr and Falk, 1999; Gneezy, 2003). We therefore expect a positive relation between x and the wage level in our gift-exchange treatment. The interesting question, however, is how the reciprocity motive interacts with the control decision. Assume that the principal pays a low wage of 10. In this case, a selfish agent chooses $x = 0$, but even a reciprocal agent will most likely choose $x < 10$. In fact, given the results from previous gift-exchange experiments, the highest x that is likely to be chosen is the one that equates the payoff of the principal and the agent, which equals $x = 20/3 < 10$ if the wage is 10. Thus, even in the presence of strongly reciprocally motivated agents, x will be higher if the principal controls than if he does not. For higher wages, however, the optimality of

---

[15] For example, one subject says, "If I feel to be forced to do something, I am no longer willing to give more than necessary. I give a little bit more than what is required, because I want to make the decision myself and without any influence from the outside."

[16] The importance of autonomy is consistent with recent evidence, e.g., from the European Survey of Working Conditions, showing that workplace flexibility has a strong positive effect on workers' job satisfaction and that this effect is mainly driven by workers' positive evaluation of an increased autonomy over how to perform their tasks and the opportunity to participate in decision-making (Thomas K. Bauer, 2004).
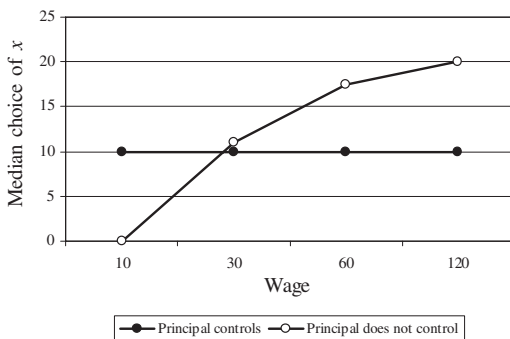
FIGURE 3. HIDDEN COSTS OF CONTROL IN THE
GIFT-EXCHANGE GAME

control versus trust depends on the interaction between perceived kindness of high wages and perceived unkindness of control. If a principal sets a high wage and refrains from controlling, he unambiguously signals his trust and the expectation of a high x. If he pays a high wage and controls, however, the signal contains, on the one hand, high expectations and kindness resulting from the high wage and, on the other hand, low expectations and distrust because of the control decision. The next result shows that reciprocal responses are in fact less pronounced if the principal controls compared to if he trusts.

RESULT 6: *We observe reciprocity in the gift-exchange treatment, i.e., a positive relation between wages and x. Reciprocity is significantly weaker, however, if the principal controls than if he does not control.*

Support for Result 6 comes from Figure 3. The figure shows, for each wage, the median values of x chosen by the agents. Both under control and under no control, wages and x-choices are positively correlated. This follows from a simple regression where we regressed performance on wages (with robust standard errors, clustered on individuals). The performance coefficient is positive and highly significant in both conditions (control condition: coeff. = 0.1909; no-control condition: coeff. = 0.2538; $p < 0.001$ in both conditions). Figure 3 reveals that the principal's controlling decision again entails a hidden cost—this time by partly crowding out agents' reciprocity. As conjectured, agents choose higher x levels for a wage of 10 if they are controlled than otherwise. For

higher wages, however, the median values of x are always higher if the principal does not control than if he does. Remember that in all our main treatments the median of x is equal to $\underline{x}$ if the principal controls. This holds also in the gift-exchange treatment. As Figure 3 shows, the median in the control condition is 10, independent of the wage, indicating that the majority of the agents choose exactly the minimum when controlled. In the no-control condition, on the other hand, median values increase in the wage, reaching a level of 20 for wages equal to 120. The resulting hidden costs are significant: if we modify agents' choice distributions such that any observation $x < \underline{x}$ is set equal to $\underline{x}$, the distributions are significantly different for each wage above the minimum wage 10 (Wilcoxon signed rank test, $p < 0.01$ for wages equal to 30, 60, or 120). Of course, there are also benefits from controlling. The selfish agents who choose $x < 10$ if not controlled are forced to choose 10 if controlled. The net effect of costs and benefits is displayed in the averages of x. For wages equal to 10, 30, 60, and 120, averages are 10.7, 12.6, 19.1, and 31.0 if the principal controls and 5.0, 10.1, 20.9, and 32.6 if the principal does not control. Thus, the hidden costs slightly outweigh the benefits for wages above 30, i.e., the negative effect of control on reciprocal agents is slightly stronger than the positive effect of control on the selfish agents.

The results from the gift-exchange treatment reveal that control partly crowds out agents' motivation to reciprocate high wages with high performance levels. Interestingly, the principals in this treatment again anticipate this. While only 10 percent of the principals trust the agent when paying a wage of 10 ($n = 10$), 50 percent of the principals trust for wages above 10 ($n = 24$). The number of principals who trust increases significantly in the wage principals pay (Spearman rank correlation, $p = 0.029$, $n = 34$). This finding also corroborates the result that there are pessimistic principals who decide to pay low wages and to control, while the optimistic principals appeal to reciprocity by paying relatively high wages and refrain from controlling.

V. Control at the Workplace: Questionnaire Data

We added some "realism" in the previous section by allowing principals to set wages. In

TABLE 4—"HOW HIGH IS YOUR WORK MOTIVATION?"

Work motivation	Scenario 1 (Supermarket)		Scenario 2 (Working times)		Scenario 3 (Job interview)		Scenario 4 (Locked door)		Scenario 5 (Internet)	
	Control	Trust	Control	Trust	Control	Trust	Control	Trust	Control	Trust
Very low	0.07	0.01	0.03	0.01	0.01	0.01	0.08	0.00	0.09	0.02
Low	0.31	0.03	0.26	0.14	0.14	0.02	0.24	0.07	0.33	0.12
Medium	0.36	0.25	0.48	0.30	0.41	0.10	0.41	0.33	0.41	0.42
High	0.23	0.60	0.20	0.45	0.39	0.53	0.25	0.50	0.15	0.42
Very high	0.03	0.11	0.03	0.10	0.05	0.34	0.02	0.10	0.02	0.02
Number of observations	199	204	204	199	203	197	199	203	203	199

this section, we go much further and study how control and explicit incentives affect motivation in typical work environments. This section illustrates the variety of applications of hidden costs of control, as well as their economic importance. Moreover, it shows that the results we obtained in stylized principal-agent relations in the lab carry over to quite realistic, everyday workplace scenarios.

Our research tool is a questionnaire in which we presented subjects with vignettes. We study five different workplace "scenarios." For each of these scenarios, we present two "conditions," one where the principal trusts the employee, and one where he controls or uses explicit incentive devices. The description of a particular scenario is identical for the two conditions. Each subject goes through all five scenarios but the subject goes through only one of the two conditions for a given scenario. We ask each subject to indicate his work motivation on a five-level scale ranging from "very low" to "very high" for each of the five scenarios/conditions.

Table 4 displays all scenarios and conditions: in the first scenario, the employee works in a supermarket and is responsible for checking the balances in the cash registers. In principle, he could easily take out money for himself but he is assumed not to do so. In the trust condition for this scenario, the principal does not check the employee; in the control condition, the principal controls whether the employee reported the cash balances truthfully. In the second scenario, the agent just started a new job and receives instructions about what he is expected to do. The agent in the trust condition is asked to meet his working time obligation exactly; in the control condition, he has to sign a formal agreement about his working time. The third scenario

is a job interview. The agent truthfully reports on his qualifications and work experience. He offers his previous employer as a reference, who can verify his statements. In the trust condition, the new employer believes what the agent says and hires him; he hires him in the control condition only after having consulted the previous employer. In the fourth scenario, the employee begins a new job in a small business, where he is told that the copying machine may not be used for private purposes. In the trust condition, the room with the copying machine is open; in the control condition, the room is locked and the employee has to ask his boss for the key. Finally, the employee in the fifth scenario works for a company that has recently provided Internet access on all personal computers, but this access may be used only for business purposes. In the trust condition, the management asks all employees to respect this rule; in the control condition, the management installs special software, which lists all Internet sites the employees have visited.

Our vignette study had 403 subjects. Thus we get a total of about 2,000 work motivation responses for the ten conditions (see Table 4). All subjects are undergraduate students from the University of Zurich. None of these subjects participated in any of our experimental treatments. The results in Table 4 resemble our findings in the experiments. Signals of distrust and control affect the agents' motivation negatively. For all work scenarios, the relative frequencies of agents indicating that they have a high or very high work motivation are always lower in the control than in the trust condition. For example, 71 percent of the subjects in the trust condition of scenario 1 indicate a high or very high work motivation. The number of

highly or very highly motivated subjects in the control condition, however, who learn that their principal controlled whether they checked the cash balances truthfully, is only 26 percent. If the principal believes the prospective employee and hires him, 87 percent of the subjects report a high or very high work motivation (scenario 3). The corresponding number drops to 44 percent if the principal first checks the reference. Scenario 4 mirrors the Hewlett-Packard example in footnote 1: work motivation is much higher if doors are open rather than locked. Likewise, formal agreements about working hours (scenario 2) or controlling employees' Internet access (scenario 5) reduces motivation. In all scenarios, control reduces work motivation significantly (Mann-Whitney test, $p < 0.0001$ for each of the five scenarios). Of course, we cannot conclude from this result that it would pay for the principal to trust his agent rather than to control him. What these results do show, however, is that there are hidden costs of control: the agent's work morale is dampened.

VI. Concluding Remarks

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. This also holds in our gift-exchange treatment, where principals can decide whether to control or trust the agent *and* determine the agent's wage payment. The reciprocal relationship between wages and effort is weaker in the presence of control in this treatment. Given the significant hidden costs of control in all our treatments, it may not come as a surprise to see that a majority of the principals in our experiment decide, in fact, to trust rather than to control their agent. The economic importance and possible applications of our experimental results are further illustrated by a questionnaire which reveals hidden costs of control in various real-life labor scenarios.

We also explore possible reasons for the existence of hidden costs of control. Agents correctly believe that principals who control expect to get less than those who don't. When asked for their emotional perception of control, most agents who react negatively say that they perceive the controlling decision as a signal of distrust and a limitation of their choice autonomy.

The main message of our paper is that control and explicit incentives 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. In fact, we show that the costs and benefits of controlling agents 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 to be suboptimal. Second, the strength of the explicit incentives is important. The results from our main treatments reveal a nonmonotonic relation between agents' performance and the strength of incentives: if the principal has only weak incentives at his disposal it is better to trust since controlling reduces motivation of the intrinsically motivated agents but increases the performance of opportunistic agents only marginally. As incentives get stronger, however, the disciplining effect eventually dominates the negative effect on motivation.

Our results suggest important implications for the design of organizations. First, the potential gains from control have to be weighted in light of the hidden costs of control (not to speak of the explicit costs of installing control devices). Faced with agents who have heterogeneous motivations, the key management goal is to develop incentive contracts that discipline the opportunistic agents without reducing the motivation of the intrinsically motivated ones. Second, when introducing incentives, the principal should be careful not to signal a negative message. Agents' performance was much better in the EX10 treatment, where control was implemented exogenously, compared to the subgame following the controlling decision of the principal. This finding corroborates, for example, the important role of third parties like consultancies, governments, etc., with regard to the implementation of control and explicit incentives. Third, our results from the gift-exchange game suggest that it is a bad decision to trust and to distrust at the same time. Paying low wages and refraining from control is certainly dominated

by low wages and control. On the other hand, when paying high wages, controlling the agent leads to a lower performance than not controlling the agent. It seems that principals have to confess their "true" expectations: either to trust or to control. Trusting a bit is likely to be interpreted as not trusting at all.

REFERENCES

▶ Aghion, Philippe, Mathias Dewatripont, and Patrick Rey. 1994. "Renegotiation Design with Unverifiable Information." *Econometrica*, 62(2): 257–82.

▶ Allen, Franklin, and Douglas Gale. 1992. "Measurement Distortion and Missing Contingencies in Optimal Contracts." *Economic Theory*, 2(1): 1–26.

Barkema, Harry G. 1995. "Do Top Managers Work Harder when They Are Monitored?" *Kyklos*, 48(1): 19–42.

Bauer, Thomas K. 2004. "High Performance Workplace Practices and Job Satisfaction: Evidence from Europe." Institute for the Study of Labor Discussion Paper 1265.

▶ Bénabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies*, 70(3): 489–520.

Bernheim, B. Douglas, and Michael D. Whinston. 1998. "Incomplete Contracts and Strategic Ambiguity." *American Economic Review*, 88(4): 902–32.

▶ Blout, Sally. 1995. "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes*, 63(2): 131–44.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166–93.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.

▶ Charness, Gary. 2004. "Attribution and Reciprocity in an Experimental Labor Market." *Journal of Labor Economics*, 22(3): 665–88.

Charness, Gary, and Martin Dufwenberg. 2004. "Promises and Partnership." Unpublished.

▶ Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117(3): 817–69.

▶ Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47(2): 268–98.

▶ Enzle, Michael E., and Sharon C. Anderson. 1993. "Surveillant Intentions and Intrinsic Motivation." *Journal of Personality and Social Psychology*, 64(2): 257–66.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2000. "Testing Theories of Fairness—Intentions Matter." Institute for Empirical Research in Economics Working Paper 63.

▶ Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2003. "On the Nature of Fair Behavior." *Economic Inquiry*, 41(1): 20–26.

▶ Falk, Armin, and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior*, 54(2): 293–315.

▶ Fehr, Ernst, and Armin Falk. 1999. "Wage Rigidity in a Competitive Incomplete Contract Market." *Journal of Political Economy*, 107(1): 106–34.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14(3): 159–81.

Fehr, Ernst, and Simon Gächter. 2002. "Do Incentive Contracts Undermine Voluntary Cooperation?" Institute for Empirical Research in Economics Working Paper 34.

▶ Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica*, 65(4): 833–60.

▶ Fehr, Ernst, George Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics*, 108(2): 437–59.

▶ Fehr, Ernst, and John A. List. 2004. "The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs." *Journal of the European Economic Association*, 2(5): 743–71.

▶ Fehr, Ernst, and Bettina Rockenbach. 2003. "Detrimental Effects of Sanctions on Human Altruism." *Nature*, 422(6928): 137–40.

▶ Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817–68.

Fischbacher, Urs. 1999. "z-Tree: A Toolbox for Readymade Economic Experiments." Institute for Empirical Research in Economics Working Paper 21.

▶Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, 6(3): 347–69.

Frey, Bruno S. 1993. "Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty." *Economic Inquiry*, 31(4): 663–70.

Frey, Bruno S. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham: Edward Elgar Publishing.

▶Frey, Bruno S., and Reto Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys*, 15(5): 589–611.

Gneezy, Uri. 2003. "Do High Wages Lead to High Profits? An Experimental Study of Reciprocity Using Real Effort." Unpublished.

Gneezy, Uri. 2004. "The W Effect of Incentives." Unpublished.

▶Gneezy, Uri, and Aldo Rustichini. 2000a. "A Fine Is a Price." *Journal of Legal Studies*, 29(1): 1–17.

▶Gneezy, Uri, and Aldo Rustichini. 2000b. "Pay Enough or Don't Pay at All." *Quarterly Journal of Economics*, 115(2): 791–810.

Ichino, Andrea, and Gerd Mühlheusser. 2004. "How Often Should You Open the Door? Optimal Monitoring to Screen Heterogeneous Agents." Institute for the Study of Labor Discussion Paper 987.

▶Lepper, Mark R., and David Greene. 1975. "Turning Play into Work: Effects of Adult Surveillance and Extrinsic Rewards on Children's Intrinsic Motivation." *Journal of Personality and Social Psychology*, 31: 479–86.

Luhmann, Niklas. 1968. *Vertrauen: Ein Mecha-nismus der Reduktion sozialer Komplexität.* 4th ed. Stuttgart: Lucius & Lucius, 2000.

▶Nöldeke, Georg, and Klaus M. Schmidt. 1995. "Option Contracts and Renegotiation: A Solution to the Hold-up Problem." *RAND Journal of Economics*, 26(2): 163–79.

Packard, David. 1995. *The HP Way: How Bill Hewlett and I Built Our Company*. New York: HarperCollins Publishers.

Pittman, Thane S., Margaret E. Davery, Kimberly A. Alafat, Kathryn V. Wetherill, and Nancy A. Kramer. 1980. "Informational versus Controlling Verbal Rewards." *Personality and Social Psychology Bulletin*, 6(2): 228–33.

▶Plant, Robert, and Richard M. Ryan. 1985. "Intrinsic Motivation and the Effects of Self-Consciousness, Self-Awareness, and Ego-Involvement: An Investigation of Internally Controlling Styles." *Journal of Personality*, 53(3): 435–49.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281–1302.

Roth, Alvin E. 1995. "Bargaining Experiments." In *The Handbook of Experimental Economics,* ed. John H. Kagel and Alvin E. Roth, 253–348. Princeton: Princeton University Press.

Sliwka, Dirk. 2003. "On the Hidden Costs of Incentive Schemes." Institute for the Study of Labor Discussion Paper 844.

▶Spier, Kathryn E. 1992. "Incomplete Contracts and Signalling." *RAND Journal of Economics*, 23(3): 432–43.