# THE CENTRE FOR MARKET AND PUBLIC ORGANISATION

The Centre for Market and Public Organisation (CMPO) is a leading research centre, combining expertise in economics, geography and law. Our objective is to study the intersection between the public and private sectors of the economy, and in particular to understand the right way to organise and deliver public services. The Centre aims to develop research, contribute to the public debate and inform policy-making.

CMPO, now an ESRC Research Centre was established in 1998 with two large grants from The Leverhulme Trust. In 2004 we were awarded ESRC Research Centre status, and CMPO now combines core funding from both the ESRC and the Trust.

Centre for Market and Public Organisation
Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
http://www.bris.ac.uk/Depts/CMPO/

*Tel: (0117) 33 10799*
*Fax: (0117) 33 10705*
*E-mail: cmpo-office@bristol.ac.uk*

# Legitimacy of Control

Wendelin Schnedler and Radovan Vadovic

August 2007

Working Paper No. 07/178

Published as IZA Discussion Paper no. 3013
(August 2007)

ISSN 1473-625X

**CMPO Working Paper Series No. 07/178**

# Legitimacy of Control

Wendelin Schnedler[1*]
and
Radovan Vadovic[2]

[1]*Department of Economics, University of Heidelberg*
[2] *Department of Economics, ITAM Mexico City*

August 2007

## Abstract
What is the motivational effect of imposing a minimum effort requirement? Agents may no longer exert voluntary effort but merely meet the requirement. Here, we examine how such *hidden costs of control* change when control is considered legitimate. We study a principal-agent model where control signals the expectations of the principal and the agent meets these expectations because he is guilt-averse. We conjecture that control is more likely to be considered legitimate i) if it is not exclusively aimed at a specific agent or ii) if it protects the endowment of the principal. Given the conjecture the model predicts that hidden costs are lower when one of the two conditions is met. We experimentally test these predictions and find them confirmed.

**Keywords:** moral-hazard, intrinsic motivation, guilt aversion

**JEL Classification:** C7, C9, M5

**Electronic version:** http://www.bris.ac.uk/Depts/CMPO/workingpapers/wp178.pdf

**Acknowledgement:** Funding by the Deutsche Forschungsgemeinschaft (SFB 504) is gratefully acknowledged.

**Address for Correspondence**
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol
BS8 1TX
Wendelin.schnedler@awi.uni-heidelberg.de*
www.bris.ac.uk/Depts/CMPO/

**CMPO** is jointly funded by the Leverhulme Trust and the ESRC

> What would, under ordinary circumstances, be justly condemned as persecution, may fall within the bounds of legitimate selfdefence.

Thomas Babington Macaulay, *The History of England*, 1849.

# 1 Introduction

Is it a good idea for a firm to impose minimum effort requirements? Controlling workers in this way has a positive *direct* effect: workers can no longer shirk. However, there may be a more subtle *indirect* effect: workers who already meet the requirement can also react to control, possibly by reducing effort. In a recent experimental study, Armin Falk and Michael Kosfeld (2006, henceforth F&K) demonstrate that the overall effect can be negative. Their result shows that subjects exert less effort when they are controlled. While F&K conclude from this finding that there are *hidden costs of control,* little is known about their extent and the circumstances under which they occur. Here, we show that hidden costs of control depend on whether control is legitimate. When it is legitimate, hidden costs are reduced.

The following example illustrates what we mean by "legitimate." According to the 2005 Internet Usage Study, about 78% of workers accessed the internet for unproductive purposes such as personal use or entertainment while at work.[1] Most employers seem to be aware of this problem and do not shy

[1] The information is from a press release of Bursteck on Business Wire Inc., March 2006.

away from monitoring the usage of internet and e-mail.[2] However, they often justify this step. As an example take the following quote from the Verizon Code of Conduct (2001):[3]

> "Sometimes it is necessary to monitor employee personal communications or computer usage or to search employee workspaces for the protection of employees, company assets and other legitimate business reasons. [The company] retains the right to monitor personal communications or computer usage or to search any and all computer property at any time, including, but not limited to, offices, desks, lockers, bags, vehicles, e-mail, voice mail, pagers, ... [company] telephone usage records and computer files."

This justification encompasses two key elements. First, control is a general policy and does not target a specific employee. Second, control prevents loss of or damage to company assets. Both issues may play a role in assuring the understanding of the worker and reducing the hidden costs of control. In line with this intuition, we put forth a legitimacy conjecture: control is more likely to be regarded as legitimate

1. if control is not aimed at a specific individual or

2. if it protects the endowment of the person who controls.

[2]The *Michigan Lawyers Weekly* reports on the 14th of August 2006 from the 2005 Electronic Monitoring & Surveillance Survey that about 75% of employers monitor employee internet connections and about 55% store and monitor their e-mail.

[3]The quote is also from the aforementioned article.

While the concept of "legitimacy" has a long tradition in political science and sociology,[4] it has, to our knowledge, not yet been studied in economics. Our legitimacy conjecture is a first step in this direction.

We provide a simple principal-agent model to formalize the consequence of this conjecture on hidden costs. The model builds on the game that F&K examine experimentally. The principal (firm, she) can decide whether to control or not. Then, the agent (employee, he) decides on effort. Control rules out effort below an exogenously given threshold level.

F&K suggest that hidden costs arise in this game because control signals a lack of trust. If the principal does not trust the agent, i.e. expects little effort of him, F&K anticipate that the agent exerts little effort as a result. We model this by using the theory of guilt-aversion proposed by Martin Dufwenberg and Gary Charness (2006) and Pierpaolo Battigalli and Martin Dufwenberg (2007). A guilt-averse agent exerts as much effort as he believes the principal expects from him because he would feel that he was letting the principal down if he exerted less effort.

The presence of the guilt-averse agent turns the game into a dynamic psychological game in which the utility of the agent depends on his own beliefs.[5] We analyze this game and find two types of equilibrium: open-door equilibria where the principal does not control because otherwise she would incur hidden costs (Proposition 1) and closed-door equilibria where

[4]For example, Max Weber (1947) describes different types of legitimacy as sources of governmental power.

[5]The theory of dynamic psychological games has been developed by Pierpaolo Battigalli and Martin Dufwenberg (2007), who build on the framework of John Geanakoplos, David Pearce, and Ennio Stacchetti (1989).

the principal controls and there are no hidden costs (Proposition 2). The type of equilibrium depends on the agent's beliefs.

The agent's beliefs also reflect whether he regards control to be legitimate or not. If control is not legitimate in the eyes of the agent, the principal cannot expect him to exert the same effort under control as he would under no control. Accordingly, the agent's belief about the principal's expectation is lower when under control than it is when under no control. On the other hand, the principal has no reason to expect less effort from the agent if the agent regards control to be legitimate. In this case, the agent's belief about the principal's expectations remains at least the same under control as under no control. In summary, the agent believes that the principal expects the same or more effort under control if (and only if) control is legitimate in the eyes of the agent. Legitimacy thus restricts beliefs. An immediate consequence of this restriction is that hidden costs occur if and only if control is not legitimate (Proposition 3).

Given this model, the legitimacy conjecture yields two testable predictions. Hidden costs of control should be smaller if control is not exclusively aimed at the agent or if control protects the principal's endowment. We test these predictions in an experiment with three treatments. In the baseline treatment, control is exclusively aimed at the agent and does not protect the principal's endowment. The second treatment differs from the baseline treatment in that control is not exclusively aimed at the subject in the role of the agent but also restricts a selfish computerized agent. The third treatment deviates from the baseline treatment in that control protects the principal's endowment. We estimate hidden costs in all treatments and find that they are significantly lower if control is also targeted at the computerized agent or if it protects the principal's endowment.

These findings contribute to the literature on the psychological effects of incentives.[6] It has been documented that material incentives may back-fire if psychological incentives are not taken into account.[7] Our findings shed light on the important question of when material incentives work and when they do not work. The model that we develop contributes to recent literature that explains why incentives may "crowd out" employee motivation. It has been suggested that incentives may be a signal from the principal to the agent about the agent's costs (Roland Bénabou and Jean Tirole 2003), about a social norm (Dirk Sliwka 2007), about the commitment of co-workers (Guido Friebel and Wendelin Schnedler 2007), or about the character of the principal (Tore Ellingsen and Magnus Johannesson 2007). The complementary explanation provided here is that incentives are informative about the expectations of the principal.

The remainder of the paper is organized as follows. In Section 2 we offer a model of hidden costs of control and in Section 3, we analyze it. Then, in Section 4, we define legitimacy and examine its consequences on hidden costs. In Section 5 we present our conjecture, derive testable predictions and propose treatments to test them. Section 6 presents our results. Finally, Section 7 summarizes our findings.

[6]This literature (for an overview see Ernst Fehr and Armin Falk 2002) incorporates a richer description of the agent's motivation in terms of equity (Ernst Fehr and Klaus Schmidt 1999) and reciprocity (Matthew Rabin 1993, Martin Dufwenberg and Georg Kirchsteiger 2004 and Armin Falk and Urs Fischbacher 2006).

[7]see for example Uri Gneezy and Aldo Rustichini (2000a, 2000b)

## 2 Modeling Hidden Costs of Control

As a starting point, consider the game from the main treatment by F&K (C10). In this game, there is a principal (she) and an agent (he). The agent is endowed with $\pi_{\mathrm{A}}^0 = 120$ points and the principal has no endowment, $\pi_{\mathrm{P}}^0 = 0$. First, the principal decides whether to impose a minimum effort requirement $\underline{x} = 10$, i.e. she can control the agent $y = \underline{x}$ or not $y = 0$. Second, the agent decides how much effort, $x$, to exert, where $x$ is an integer between $y$ and the endowment of 120. The strategy of the agent is a pair $(x^{\mathrm{c}}, x^{\mathrm{nc}})$ that specifies an effort choice when the principal controls and when she does not control. Each unit of effort costs the agent $x$ and increases the payoff of the principal by $2 \cdot x$. Hence there are efficiency gains if the agent exerts effort. The monetary payoffs are:

$$\pi_{\mathrm{A}}^1(y, x) = \pi_{\mathrm{A}}^0 - x \text{ for the agent and} \tag{1}$$

$$\pi_{\mathrm{P}}^1(y, x) = \pi_{\mathrm{P}}^0 + 2 \cdot x \text{ for the principal.} \tag{2}$$

We can distinguish two effects of control:

**Definition 1 (Direct and indirect effect of control)** *The direct effect of control forces the agent to supply at least the required benchmark,* $\underline{x}$*. The indirect effect of control reflects any other response by the agent.*[8]

The indirect effect captures the consequences on the principal's payoff due to the agent's psychological reaction to control. A particular interesting case is when this effect is negative.

[8]Mathematically, the direct effect on the principal's payoff amounts to $2 \cdot (\underline{x} - x^{\mathrm{nc}})$ if $x^{\mathrm{nc}} < \underline{x}$ and zero else. The indirect effect is the overall effect, $2 \cdot (x^{\mathrm{c}} - x^{\mathrm{nc}})$, minus the direct effect. It can be more succinctly written as $2 \cdot (x^{\mathrm{c}} - \max\{\underline{x}, x^{\mathrm{nc}}\})$.

**Definition 2 (Hidden costs of control)** *There are hidden costs of control if and only if the indirect effect of control is negative.*

Later, we estimate direct, indirect, and overall effect using the data from our treatments and test for the presence of hidden costs.

In light of these definitions, the finding by F&K that control leads to lower effort has three implications. First, there is an indirect effect of control. Second, this effect is negative, i.e. there are hidden costs of control; and third, it outweighs the direct effect of control.

F&K suggest that hidden costs arise because control signals a lack of trust. If the principal does not trust the agent, i.e. expects little effort of him, the agent exerts little effort. The idea that control signals distrust can be formalized using the theory of guilt-aversion.[9] A guilt-averse agent exerts as much effort as the principal expects since otherwise he would suffer from feeling guilty.

Suppose the principal expects the agent to supply a certain level of effort when she controls and a possibly different level when she does not control. In other words, the principal has some first-order belief about effort which depends on control. Further, suppose that the agent has a second-order belief:[10] he has an estimate of the principal's expectations given that the principal controls, $\mu^{\mathrm{c}}$, and given that the principal does not control, $\mu^{\mathrm{nc}}$.

The utility of a guilt-averse agent depends not only on monetary payoffs but also on his belief about what the principal expects him to do. Formally,

[9]For a formal and more complete introduction to guilt-aversion see the articles by Martin Dufwenberg and Gary Charness 2006 and Pierpaolo Battigalli and Martin Dufwenberg, 2007)

[10]Higher order beliefs could be defined accordingly but are not needed for our analysis.

we denote the guilt felt by the agent when he thinks the principal expects $\mu$ and he supplies $x$ with $G(\mu, x) := \max\{0, 2 \cdot (\mu - x)\}$. Let $\theta \geq 1/2$ describe the degree of guilt-aversion of the agent,[11] then the overall utility for the agent from monetary payoff and guilt is:

$$u_{\mathrm{A}}\left(y, (x^{\mathrm{nc}}, x^{\mathrm{c}}) \mid (\mu^{\mathrm{c}}, \mu^{\mathrm{nc}})\right) \quad = \quad \pi_{\mathrm{A}}^{1}\left(y, x^{i}\right) - \theta \cdot G(\mu^{i}, x^{i}), \tag{3}$$

where $i = \mathrm{c}$ if the principal controls and $i = \mathrm{nc}$ else. For simplicity, we assume that the principal only cares about her monetary payoffs.

# 3 Analysis

The introduction of the guilt-averse agent leads to two equilibria (respective proofs are in Appendix A.2).

**Proposition 1 (Open-door equilibrium)** *There is an open-door equilibrium where the principal does not control ($y = 0$) and the agent's effort matches the agent's beliefs ($x^{nc} = \mu^{nc}, x^{c} = \mu^{c}$) if and only if the agent's beliefs satisfy:*

$$\mu^{nc} - \mu^{c} \geq 0. \tag{4}$$

As the agent matches expectations, effort under no control is always larger than or equal to effort under control. In the open-door equilibrium, the indirect effect of control may hence be negative and control can entail hidden costs.

[11] If $\theta < 1/2$ marginal costs of exerting effort always exceed the marginal guilt and guilt-averse agents act like selfish agents.

**Proposition 2 (Closed-door equilibrium)** *There is a closed-door equilibrium where the principal controls ($y = \underline{x}$) and the agent's effort matches the agent's beliefs ($x^{nc} = \mu^{nc}, x^c = \mu^c$) if and only if the agent's beliefs satisfy:*

$$\mu^{nc} - \mu^c \leq 0. \tag{5}$$

In this second equilibrium, the indirect effect cannot be negative and control does not lead to hidden costs. Which equilibrium arises depends on the initial beliefs of principal and agent. In the next section, we examine these beliefs and their relation to legitimacy more closely.

# 4 Legitimacy

Legitimacy itself is a broad concept. The Oxford English Dictionary lists twelve different definitions of "legitimate."[12] Here, we focus on one particular but crucial aspect of legitimacy: its consequence on the agent's beliefs.

Consider a guilt-averse agent who provides more than the required effort in the absence of control: $x^{\mathrm{nc}} > \underline{x}$. This agent is not affected directly by control. However, he could be affected indirectly. On the one hand, if the agent does not consider control to be legitimate, the principal cannot expect his effort under control to match or exceed the effort when she does not control. Accordingly, the agent's beliefs about the principal's expectations are lower under control: $\mu^{\mathrm{nc}} > \mu^{\mathrm{c}}$. On the other hand, if the agent regards control to be legitimate, the principal can expect to get at least the same

[12]The most relevant definition for our purposes is in line with the initial quote from Macaulay and reads "conformable to law or rule." Oxford English Dictionary, 2nd Edition, 1989, Oxford: Oxford University Press.

effort under control as under no control and the agent believes that the principal's expectations are not lower under control: $\mu^{\mathrm{nc}} \leq \mu^{\mathrm{c}}$. The following definition reflects this intuition.

**Definition 3 (Legitimate Control)** *Control is regarded to be legitimate by the agent if and only if his beliefs about the principal's expectation are at least as high under control as when under no control:* $\mu^{c} \geq \mu^{nc}$.

In psychological games, the source of initial beliefs is un-modeled. Our definition of legitimacy structures these beliefs and is closely related to hidden costs of control.

**Proposition 3 (Legitimacy and Hidden Costs)** *There are hidden costs of control if and only if the agent does not regard control as legitimate.*

The proof follows immediately from the observation that the agent's actions meet the principal's expectations (see Lemma 1 in the appendix). It is thus valid as long as the principal has the choice between control and no control and the agent's utility and cost function are as previously specified.

If control is not legitimate then the closed-door equilibrium does not exist, we are only left with the open-door equilibrium, and control entails hidden costs. However, we have not yet addressed the question under what circumstances the agent regards control to be legitimate and will do so in the following section.

## 5 Hypotheses and Treatments

Whether an individual regards control to be legitimate may vary between individuals and depend on the context in which control is applied. Our starting point is the following conjecture.

**Conjecture 1 (Legitimacy Conjecture)** *Control is more likely to be considered legitimate*

1. *if it is not exclusively aimed at a particular agent or*

2. *if it protects the endowment of the principal.*

Within our model, the legitimacy conjecture yields two testable predictions: hidden costs of control should be lower if control is not targeted at a specific agent or if it protects the principal against theft. Next, we suggest two variations of the aforementioned game in order to test these predictions.

For comparison, we replicate the C10 treatment by F&K as our BASELINE treatment. Our second treatment (ROBOT) varies the BASELINE in that the principal is no longer certain to be matched with a human subject. With a probability of one half, she is matched with a computerized agent that always chooses the smallest possible effort, i.e. $x = 0$ under no control and $x = 10$ under control. Accordingly, the payoffs to the principal are 0 under no control and 20 under control. If the principal interacts with the computerized agent, the human agent is not affected by this and keeps his endowment. When deciding on control, the principal does not know whether she is matched with the human or with the computerized agent. Control is hence not exclusively aimed at the human agent. According to our conjecture, more subjects should consider control to be legitimate in the ROBOT treatment than in the BASELINE treatment. Applying Proposition 3, we obtain the following testable hypothesis.

**Hypothesis 1** *Control leads to lower hidden costs in the ROBOT than in the BASELINE treatment.*

Our third treatment (ENDOWMENT) differs from the BASELINE treatment in that the agent has the possibility to take from the principal. The only difference is that effort is labeled differently. While in the BASELINE treatment effort ranges from zero to 120 and the minimum requirement is 10, effort in the ENDOWMENT treatment is between $-10$ and 110 and the minimum requirement is 0. If the agent neither takes nor gives, the principal has an endowment of $\pi_{\mathrm{P}}^{0} = 20$ and the agent of $\pi_{\mathrm{A}}^{0} = 110$. In the ENDOWMENT treatment, control protects the principal's initial endowment. So, if the legitimacy conjecture holds, Proposition 3 implies the following hypothesis.

**Hypothesis 2** *Control leads to lower hidden costs in the ENDOWMENT than in the BASELINE treatment.*

In order to test our hypotheses, we ran a total of 12 sessions: 4 for each treatment. All sessions were conducted in the experimental laboratory at the University of Mannheim. Subjects were primarily undergraduate students who were randomly recruited from a pool of approximately 1000 subjects using an e-mail recruitment system. Each subject only participated in one of the treatments. The software was written in Visual Basic 6 and the experiment lasted approximately 60 minutes (including time for reading the instructions and receiving payments).

After the subjects arrived at the laboratory, they were randomly and anonymously matched in pairs and seated at the computer terminals. They were handed instructions (included in the appendix) and given 15 to 20 minutes to study them. After everyone had finished reading, they were asked to complete a series of questions designed to verify their understanding of the experiment. Once all questions had been answered successfully, the experiment began. The game was not repeated: each subject only had to make one decision. Subjects in the role of the principal had to decide whether to control or not. For subjects in the role of the agent, we used the strategy method to elicit their decisions: the subjects had to decide on the effort level under control and no control. The strategy method is used because it allows us to identify hidden costs of control even if the behavior of principals shows little variation. For example, we learn what agents would have done under no control even if all principals control. At the end of the experiment, we paid each subject privately in cash. All material payoffs were explained in points that were converted using the rate that 1 point=10 cents. Subjects received a show-up fee of €4 and their total earnings were on average about €10 for the whole experiment.

# 6 Results

Our first finding replicates an important aspect of the work by Falk and Kosfeld (2006).

**Result 1** *In the BASELINE treatment, control entails hidden costs.*

The left panel in Figure 1 shows the distribution of effort choices under control and no control. The distribution under control is stochastically smaller. This indicates that a sizeable fraction of subjects provides more effort in the absence of control. The right panel in the same figure depicts the distribution of the individual difference between the effort under control and without control for each subject. It shows that around 37% of the subjects

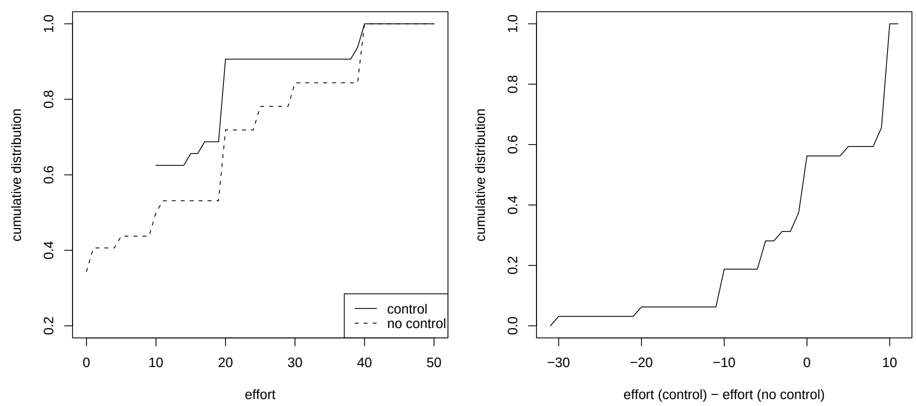


Figure 1: Cumulative distribution of effort (left panel) and of within-subject differences in effort choices for the BASELINE treatment

give less when they are controlled and 18% reduce their effort by ten points or more.

When we estimate the indirect effect of control on effort,[13] we find that it amounts to −6.88 points. On average, principals thus incur hidden costs

[13]The indirect effect is estimated by replacing $x^{c}$ and $\max\{x^{nc}, \underline{x}\}$ with the respective sample means of the variables. We test whether there are hidden costs by checking whether the median of the distribution of $x^{c}$ is smaller than the median of $\max\{x^{nc}, \underline{x}\}$. F&K do not explicitly define "hidden costs" but use the same testing procedure to test whether control has a "behavioral impact."

of about 7 points when controlling. Is this loss statistically significant? In line with Falk and Kosfeld (2006), we find that hidden costs of control are significant (two-sided exact Wilcoxon signed rank test has a p-value below 0.001). In the BASELINE treatment, the direct effect amounts to 8.31 points so that the overall effect is $8.31 - 6.88 = 1.43$; we cannot reject the hypothesis that the overall effect is zero (The two-sided Wilcoxon signed rank test has a p-value of 0.38). So, we cannot replicate the finding of F&K that the principal is worse off if she controls.

After having established the existence of costs of control, we now address our main hypotheses by examining how these costs change across treatments.

**Result 2** *Hidden costs of control are significantly lower in the ROBOT than in the BASELINE treatment.*

In the ROBOT treatment one-third of the subjects voluntarily exert an effort above the minimum requirement even if they are not controlled. The left panel in Figure 2 illustrates that agents still choose higher effort in the absence of control, while the right panel indicates that the differences between effort under control and in the absence of control are considerably larger in the ROBOT treatment than in the BASELINE treatment. In the ROBOT treatment, only 9% of the subjects choose to exert less effort under control while about 37% do so in the BASELINE treatment. Likewise, the share of subjects who "punish" control by reducing effort by 10 or more points drops from 18% in the BASELINE to only 6% in the ROBOT treatment. The estimate for the indirect effect of control amounts to $-2.06$ points. While control still leads to hidden costs, they are no longer significantly different

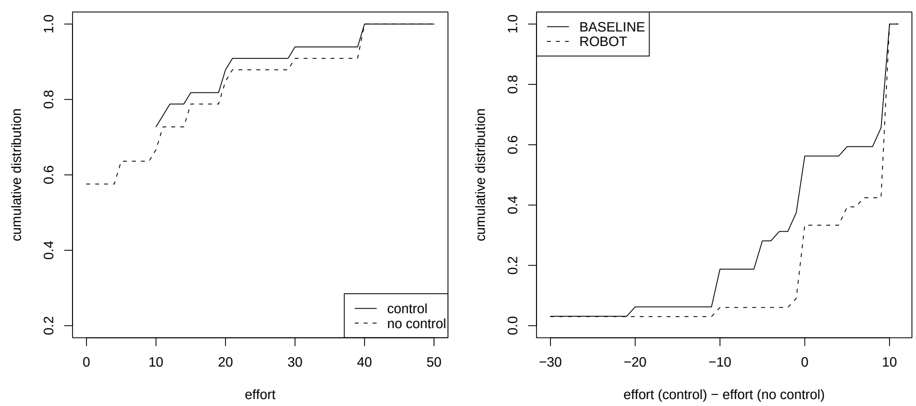


Figure 2: Cumulative distribution of effort for the ROBOT treatment (left panel) and of the differences in effort choices for each subject in the BASELINE and ROBOT treatment (right panel)

from zero (the one-sided Wilcoxon signed rank test has p-value of 0.78). Hidden costs in the ROBOT treatment are less than a third of those in the BASELINE treatment. We can verify whether this drop is significant by testing whether the indirect effect in the ROBOT treatment is smaller than or equal to that in the BASELINE treatment.[14] Based on the two-sample one-sided test, this hypothesis is rejected (the Wilcoxon signed rank test has a p-value below 0.01). The hidden costs of control thus decrease when moving to an environment where control is not specifically aimed at the agent. Next, we turn to the ENDOWMENT treatment.

**Result 3** *Hidden costs of control are significantly lower in the ENDOWMENT than in the BASELINE treatment.*

[14] Recall that a smaller indirect effect means larger hidden costs.

One-third of the subjects voluntarily give more than the required number of points. Subjects in the role of the agent, however, do no longer choose a different effort under control and in the absence of control: the left panel in Figure 3 shows that the distribution of effort is not stochastically smaller when the agent is controlled. The share of subjects that "punish" the principal for controlling is lower than in the BASELINE treatment and their reduction in effort is smaller too: only 11% exert less effort when being controlled and only about 5% reduce their effort by more than 10—see right panel in Figure 3. Recall that the indirect effect can be positive if agents

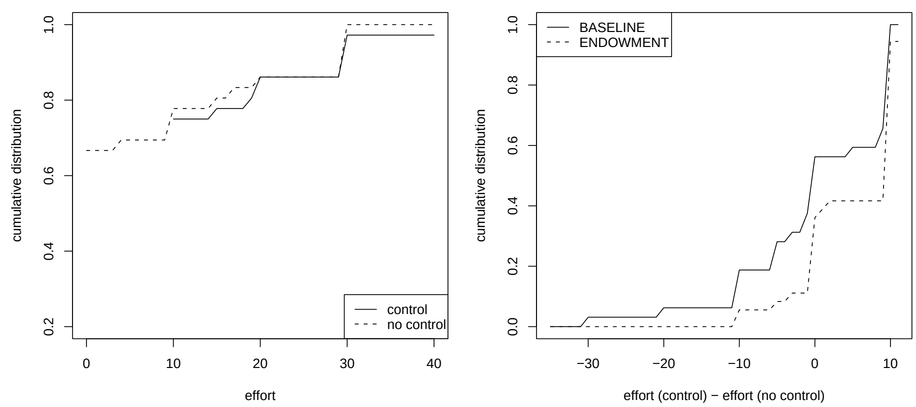


Figure 3: Cumulative distribution of effort (left panel) and of within-subject differences in effort choices for the ENDOWMENT treatment

choose to exert more effort when they are controlled. If the indirect effect is positive, then there are, of course, no hidden costs of control. This is indeed the case for the ENDOWMENT treatment: the estimate for the indirect effect amounts to 6.11 points. If we use the same testing approach as before, we find that the indirect effect in the ENDOWMENT treatment (and hence the hidden costs of control) is significantly lower than in the BASELINE treatment (Wilcoxon signed rank test has a p-value below 0.01).[15]

The findings from both treatments mesh well with the idea of legitimacy: if control is legitimate (according to our definition) then principals are not punished for controlling. Are principals aware of this?

**Result 4** *In the ROBOT treatment and in the ENDOWMENT treatment, principals control significantly more often than in the BASELINE treatment.*

In the BASELINE treatment, the principal controls in 23 of 33 cases. These numbers contrast with the ROBOT treatment, where she controls in 32 out of 33 cases, and the ENDOWMENT treatment, where control happens in 32 out of 36 cases. Figure 4 depicts the shares of principals that control for the three treatments. The share in BASELINE is significantly lower than in ROBOT (p-value for Pearson's $\chi^2$-test: $< 0.01$) and in ENDOWMENT (p-value for Pearson's $\chi^2$-test: 0.046). Principals seem to be aware of the fact that controlling leads to lower costs when control is legitimate.

When interpreting this result, it is important to note that control is attractive in all three treatments: given the actual behavior of agents, control is an optimal strategy.[16] So a significantly larger proportion of principals controls in the ROBOT treatment *although* control already maximizes surplus in

[15] The positive sign of the indirect effect is due to a single subject who exerts the maximal effort when being controlled and no effort otherwise. Excluding this agent from the data does not change the results: although the indirect effect becomes negative, hidden costs remain significantly lower in the ENDOWMENT treatment, and the Wilcoxon signed rank test still has a p-value below 0.01.

[16] Based on the C10 treatment by F&K, we did not expect control to be optimal in the BASELINE treatment.

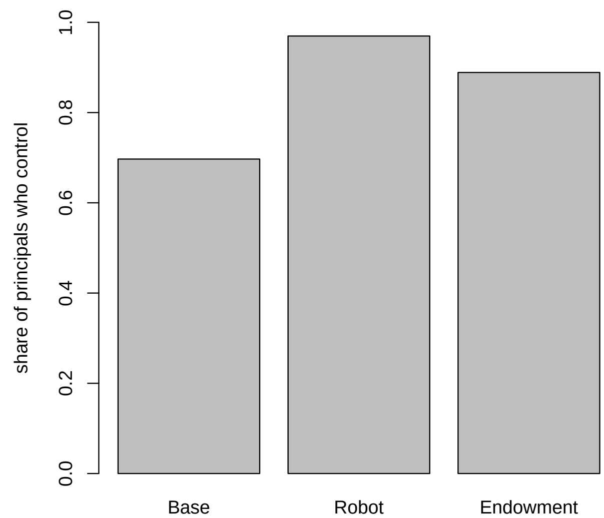


Figure 4: Control by principals across treatments

the BASELINE treatment. Moreover, more agents anticipate that principals control in the ROBOT than in the BASELINE treatment:[17] in the BASELINE treatment 28 of 32 agents expect the principal to control while all 33 agents in the ROBOT treatment expect to be controlled; these differences are weakly significant (p-value for Pearson's $\chi^2$-test: 0.06). Likewise more agents believe that principals control in the ENDOWMENT treatment although the difference is not significant. In the ROBOT and ENDOWMENT treatment, control is anticipated by the agent and is hence less informative as a signal of expectations. This may explain why the response of agents to control is weaker in these treatments than in the BASELINE treatment. We take this as additional, albeit anecdotal, confirmation that control acts as a signal of distrust.

[17]We asked agents what they believed the principal would do. The question was not salient.

# 7 Conclusion

Control can have two types of consequences: the desired direct effect of preventing a certain behavior and the indirect effect of provoking counter-productive reactions. We offer a model to examine the interplay of these two effects. Based on this model, we obtain the predictions that control provokes less detrimental behavior if it is not aimed at a specific person or if it protects property. We test these predictions in an experiment. The predictions are borne out by the data. Our results highlight the fact that the behavioral consequences of control are specific to the context in which control is used. Armin Falk and Michael Kosfeld (2006) suggest that control may be an attractive option for the principal only when it is particularly effective, in other words, when principals can enforce a high minimum effort level. In this case, the direct, disciplining effect of control dominates any other indirect, behavioral response to control. By looking at the two effects separately, our study uncovers another reason for principals to control: the detrimental behavioral response to control (the hidden costs) may vanish. Our experiment demonstrates that hidden costs are minimal if agents do not take control personally or if they are able to relate to the principal's decision to protect her property.

# References

**Battigalli, Pierpaolo and Martin Dufwenberg**, "Guilt in Games," *American Economic Review*, May 2007, *97* (2), 170–176. Papers and Proceedings.

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

**Charness, Gary and Martin Dufwenberg**, "Promises and Partnership," *Econometrica*, 2006, *74*, 1579–1601.

**Dufwenberg, Martin and Georg Kirchsteiger**, "A theory of sequential reciprocity," *Games and Economic Behavior*, May 2004, *47* (2), 268–298.

**Ellingsen, Tore and Magnus Johannesson**, "Pride and Prejudice: The Human Side of Incentive Theory," 2007. mimeo, Stockholm School of Economics.

**Falk, Armin and Michael Kosfeld**, "The Hidden Costs of Control," *American Economic Review*, 2006, *96* (5), 1611–1630.

_ **and Urs Fischbacher**, "A theory of reciprocity," *Games and Economic Behavior*, February 2006, *54* (2), 293–315.

**Fehr, Ernst and Armin Falk**, "Psychological Foundations of Incentives," *European Economic Review*, 2002, *46*, 687–724.

_ **and Klaus M. Schmidt**, "A Theory Of Fairness, Competition, And Cooperation," *The Quarterly Journal of Economics*, August 1999, *114* (3), 817–868.

**Friebel, Guido and Wendelin Schnedler**, "Team Governance: Empowerment or Hierarchical Control," 2007. mimeo, University of Heidelberg.

**Geanakoplos, John, David Pearce, and Ennio Stacchetti**, "Psychological Games and Sequential Rationality," *Games and Economic Behaviour*, March 1989, *1* (1), 60–79.

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

_ **and** _ , "Pay Enough or Don't Pay at All," *The Quarterly Journal of Economics*, August 2000, *115* (3), 791–810.

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

**Sliwka, Dirk**, "On the Hidden Costs of Incentive Schemes," *American Economic Review*, 2007, *97* (3), 999–1012.

**Weber, Max**, *The Theory of Social and Economic Organization*, Glencoe: The Free Press and the Falcon's Bring Press, 1947. translated by A.M. Henderson and Talcott Parsons.

# A Appendix

## A.1 Auxiliary results

**Lemma 1** *In all equilibria, the agent meets the expectations of the principal.*

**Proof.** Consistency requires that the expected values of the agent's second order beliefs are identical to that of the principal's first order beliefs. The principal's expectations are thus $\mu^{\text{nc}}, \mu^{\text{c}}$ in equilibrium. Suppose the agent chooses a strategy such that $\tilde{x}^i > \mu^i$ instead of $x^i = \mu^i$. Then, effort costs increase by $\tilde{x}^i - x^i$ and the deviation is not profitable. On the other hand, if the agent chooses a strategy $\tilde{x}^i < \mu^i$ instead of $x^i = \mu^i$, then guilt increases by more than $\tilde{x}^i - x^i$ and the deviation is also not worthwhile. Thus, the agent chooses $x^i = \mu^i$ and meets the expectations. ■

## A.2 Proof of open- and closed-door equilibria

We present the result and the proof for a more general model. Denote by $x_{\mathrm{L}}$ the lowest possible effort if the principal does not control. Suppose the principal is only matched with the agent with probability $p$, receives $y$ with probability $1-p$, and does not know whether she is matched with the agent while deciding whether to control ($y = \underline{x}$) or not ($y = x_{\mathrm{L}}$).

**Proposition 4 (Equilibria)** *There is an open-door equilibrium where the principal does not control ($y = x_L$) and the agent's effort matches the agent's beliefs ($x^{nc} = \mu^{nc}, x^c = \mu^c$) if and only if the agent's beliefs satisfy:*

$$\mu^{nc} - \mu^c \geq \frac{1-p}{p}(\underline{x} - x_L). \tag{6}$$

*There is a closed-door equilibrium where the principal controls ($y = \underline{x}$) and the agent's effort matches the agent's beliefs ($x^{nc} = \mu^{nc}, x^c = \mu^c$) if and only if the agent's beliefs satisfy:*

$$\mu^{nc} - \mu^c \leq \frac{1-p}{p}(\underline{x} - x_L). \tag{7}$$

**Proof.** In any equilibrium, the agent and the principal have correct beliefs. By Lemma 1 the agent meets expectations, has no profitable deviation, and plays $x^{\mathrm{nc}} = \mu^{\mathrm{nc}}$ and $x^{\mathrm{c}} = \mu^{\mathrm{c}}$. Let us now turn to the principal. The payoff of the principal if she controls is: $2 \cdot (x^{\mathrm{c}} p + (1-\underline{x})(1-p))$. Under no control, it amounts to: $2 \cdot (x^{\mathrm{nc}} p + (1-x_{\mathrm{L}})(1-p))$. Accordingly, a deviation to control in an open-door equilibrium is not profitable if $\mu^{\mathrm{nc}} - \mu^{\mathrm{c}} \geq \frac{1-p}{p}(\underline{x} - x_{\mathrm{L}})$. Analogously, a deviation to 'no control' in a closed-door equilibrium is not profitable if $\mu^{\mathrm{nc}} - \mu^{\mathrm{c}} \leq \frac{1-p}{p}(\underline{x} - x_{\mathrm{L}})$. ∎

# B Instructions for the ROBOT treatment

## General Instructions for Participants

You are now participating in an economic experiment. Please read the following instructions carefully. The instructions will provide you with all the information you require for participation in the experiment. Please ask for assistance if there is something that you do not understand. Your question will be answered at your workplace. There is a strict prohibition of communication during the experiment.

You will receive a show-up fee of 4 Euro at the beginning of the experiment. Over the course of the experiment you can earn points. This income will be converted into Euro at the end of the experiment. Please note that:

1 point = 10 Cent

The converted income and the show-up fee will be paid out in cash at the end of the experiment.

## The Experiment

In this experiment, each participant A is associated with a participant B in a group of two. No participant knows with whom he is associated; all decisions are made anonymously.

At the beginning of the experiment, Participant A receives 120 points as an endowment and participant B 0 points.

## Decision of Participant A

Participant A can decide how many points he wants to give to participant B. These points are deducted from A's endowment, doubled and added to B's endowment. Each point that A gives to B thus reduces the income of A by one point and increases that of B by two points.

After participant A has chosen, the incomes hence amount to:

| | |
|---|---|
| Income of Participant A: | 120 – chosen points |
| Income of Participant B: | 0 + 2 * chosen points |

The following examples illlustrate the computation of the income:

Example 1: A gives 0 points (chosen points: 0). Then, A's income is 120 and B's 0 points.
Example 2: A gives 20 points (chosen points: 20). The income is 100 for A and 40 for B.
Example 3: A gives 80 points (chosen points:80). The income is 40 for A and 160 for B.

## Decision of Participant B

Participant B may require to be given at least 10 points or he can leave the choice of points completely free. There are hence two cases:

**Case 1**: Participant B requires to be given at least 10 points. Then participant A can chose any (integer) amount **between 10 and 120** points. In this case the income of B is at least 20 points.

**Case 2**: Participant B leaves the choice of points free. Then, participant A can chose any (integer) amount **between 0 and 120** points. In this case the income of B can amount to 0 points.

## Decision of Computer Program

A random draw decides whether the decision of participant A has any consequences at all. With a probability of 50%, participant A is replaced by a computer program. In this case, participant A's income is equal to his endowment and the income of participant B results from the number of points chosen by the computer program. The program always chooses the *smallest possible* amount of points, i.e. 10 points if participant B requires at least 10 points (case 1) and 0 points if B leaves the choice free (case 2).

Suppose that the program replaces participant A. If B leaves the choice free (case 2), then the program will give B the lowest possible amount, i.e. 0 points. The income of B is then 0 points. If B requires to be given at least 10 points (case 1), the program gives 10 points. In this case, participant B gets at least 20 points.

When participant B decides whether to leave the choice free or not, he does not know whether he faces participant A or the program.

## The Stages of the Experiment

1. Participant B decides whether to require at least 10 points or leave the choice free.

2. Participant A determines a number of points. If participant B requires at least 10 points (case 1), participant A can select a number of points between 10 and 120. If participant B leaves the choice free (case 2), participant A can select a number of points between 0 and 120.

3. A random move decides whether participant A or a program determines the income. There is a probability of 50% that A's number of points determines the size of the participants' income. There is a probability of 50% that A keeps his initial endowment and the program determine's B's income. If participant B requires at least 10 points (case 1), the program gives 10 points and B's income is 20 points. If participant B leaves the choice free (case 2), the program gives 0 points and B's income amounts to 0 points.

4. Participant A and participant B each learn the decision of the other, the result of the draw, and the size of the resulting incomes.

Then, the experiment is over.