Hidden Costs of Control and Workplace Arrangements: Experimental Evidence from the Internet and the Laboratory*

Katrin Schmelz† and Anthony Ziegelmeyer‡

This version: July 2016

Abstract

This paper reports an experiment designed to assess the influence of workplace arrangements on the magnitude of hidden costs of control. We compare behavior in an Internet and a laboratory principal-agent game where the principal can control the agent by implementing a minimum effort requirement before the agent chooses an effort costly to her but beneficial to the principal. We employ the same standard subject pool of students and identical experimental procedures in both treatments. Still, online subjects enjoy greater anonymity than lab subjects, they interact in a looser environment than the laboratory, and there is a larger physically-oriented social distance between them. Differences between the online and lab interaction environments capture essential differences between working from home and working at the office arrangements. We find that hidden costs of control are significantly lower in the Internet than in the laboratory treatment. The treatment effect is driven by both higher intrinsic motivation and stronger control aversion in the laboratory. Agents' effort differences are fairly stable over time in both treatments which indicates that even experienced agents react less negatively to control online than in the laboratory.

Keywords: Hidden costs of control; Laboratory; Internet; Workplace arrangements.

JEL Classification: C81; C90; C93; M52.

1 Introduction

Recent data from the American Time Use Survey show that the share of U.S. workers doing some or all of their work at home grew from 19 percent in 2003 to 24 percent in 2015.[1] A similar trend is observed in Europe with a quarter of the employees mostly working from places other than the office in 2010 (Eurofound, 2012). Like other flexible scheduling and work arrangements, working from home (WFH) challenges existing managerial approaches designed for office employees since such approaches might be

*We are particularly grateful to Urs Fischbacher for his invaluable advice in numerous fruitful discussions, and we appreciate Simon Gächter's stimulating feedback. We would like to thank Irenaeus Wolff and the lively research group at the Thurgau Institute of Economics (TWI) for many helpful comments, and we appreciate inspiring discussions with participants at the ESA European Meeting 2015 in Heidelberg. This study could not have been conducted without the invaluable efforts of Sirko Schindler, Susanne Schmidt and Andreas Lehmann, who did most of the programming on the internet platform. Numerous student assistants provided amazing research assistance. Representative we thank Evgenia Grishina, Christin Schulze and Jonas Gehrlein for testing the internet platform again and again. We acknowledge financial support by the Max Planck Society, the Ministry of Science, Research and the Arts of the state of Baden-Württemberg, and the Thurgau Institute of Economics.

†Corresponding author: University of Konstanz, PO Box D 131, D-78457 Konstanz, Germany, and Thurgau Institute of Economics, Hauptstrasse 90, CH-8280, Kreuzlingen, Switzerland. Email: katrin.schmelz@uni-konstanz.de.

‡Queen's Management School, Queen's University Belfast, Northern Ireland, United Kingdom. Email: email@anthonyziegelmeyer.com.

[1] The American Time Use Survey, which is conducted by the U.S. Census Bureau for the Bureau of Labor Statistics, is a continuous survey about how individuals age 15 and over spend their time.

inadequate to supervise and elicit performance from distant employees. To deal with the lack of direct oversight, employers could either develop supervisory relations based on trust and autonomy or they could turn to tougher supervisory procedures. Whether WFH calls for a different managerial approach heavily depends on whether the nature of the employment relationship, close or distant, influences work motivation or employees' reactions towards supervision.

Bruno S. Frey and coauthors have repeatedly argued that the closer the relationship between employers and employees the more likely controlling reduces work effort and performance (Frey, 1993, 1997; Frey and Jegen, 2001). Close employment relationships foster the intrinsic motivation of employees, and as a consequence an intervention of the employer perceived to be controlling is likely to crowd out work motivation. In distant employment relationships intrinsic motivation is less present and crowding out plays little role. The imposition of tougher controlling on distant employees is therefore less likely to backfire on employers. Frey's hypothesis suggests that employers should control office and WFH employees differently and adjust to the lack of day-to-day personal oversight in WFH arrangements with new ways of tightly controlling work. This perspective contrasts with the advice commonly offered by management consultants to supervisors of WFH employees according to which the management of these employees is best secured by an emphasis on trust rather than close regulation. However, interview-based evidence has questioned this oversimplified recommendation (Felstead, Jewson, and Walters, 2003). At the present time, researchers and practitioners are still debating about the most effective way to manage distant employees (see Lautsch, Kossek, and Eaton, 2009, and references therein).

This paper reports an experiment designed to assess the influence of workplace arrangements on the magnitude of hidden costs of control. We compare behavior in an Internet and a laboratory implementation of a principal-agent game where the principal can control the agent by imposing either a low or a medium effort level before the agent chooses an effort costly to her but beneficial to the principal. Our principal-agent game is a straightforward extension of the laboratory game used by Falk and Kosfeld (2006) in their main treatments. The authors show that principals suffer *hidden costs of control* because some of the agents reduce their work motivation when facing managerial control and that these agents perceive the controlling decision as a signal of distrust or a limitation of their choice autonomy. Our laboratory and Internet interaction environments differ in three aspects which are likely to influence the behavior of subjects. First, because they are recruited to a single location our lab subjects see each other before entering the agency relationship while our online subjects are paired across different locations and therefore enjoy greater subject-subject anonymity.[2] Accordingly, lab subjects might consider the possibility of running into their counterparts after the session and engage in reputation-building behavior (Eckel and Wilson, 2006).[3] This could foster the intrinsic motivation of lab agents to perform in the interest of principals and increase their negative reactions to control, but might also lead lab principals to trust their agents. Second, nearly all aspects of the laboratory setting, including physical features and even the possibility of getting up and doing something else, are determined by the experimenter whereas the Internet setting offers many more alternative activities and distractions as online subjects make their interactive choices at their place of choice (most likely home). The fact that online subjects complete their tasks in a less constrained environment than lab subjects implies that online agents have higher autonomy outside of the agency relationship which might lead them to accept more easily a reduced choice autonomy in

[2] Additionally, lab subjects might feel scrutinized by the experimenter which could lead lab agents to react more negatively to managerial control than online agents. However, we believe that our experimental procedures induced a rather comparable level of experimenter-subject anonymity in both treatments especially as subjects always collected their earnings in the laboratory. Note also that Barmettler, Fehr, and Zehnder (2012) fail to confirm that the absence of experimenter-subject anonymity in laboratory experiments leads to an overestimation of the importance of social preferences.

[3] Still, as is common practice in economic experiments, lab subjects were seated in private cubicles, they gave responses anonymously and they were paid in private.

the agency relationship. Third, online principal-agent pairs are physically disconnected contrary to lab pairs which interact in the same location. Charness, Haruvy, and Sonsino (2007) explore the effect of increasing the physically-oriented social distance between subjects and find significantly less influence for social preferences online than in the laboratory. This evidence suggests that we might observe higher levels of managerial control, less work motivation and lower negative reactions to control in the Internet treatment than in the laboratory treatment. The three distinctive features of our Internet interaction environment are inextricable components of WFH arrangements and we did not design our experiment to assess their relative impact on behavior in the agency relationship—exploring this issue is left for future work. Hereafter, we simply refer to agency relationships as being more distant over the Internet than in the laboratory which encompasses greater subject-subject anonymity, a less constrained environment, and a larger physically-oriented social distance.

We acknowledge that our experiment uses a weak manipulation of the distance in agency relationships. Employers perceive office employees closer to them than WFH employees not only because of their physical proximity per se but also because this physical proximity leads to more frequent interactions which in turn might generate some emotional proximity. In this sense, close and distant agency relationships in Frey's hypothesis might be better understood as interpersonal and impersonal agency relationships. Our weak manipulation is, however, a useful starting point since additional aspects of the agency relationship, like emotional proximity, can easily be incorporated into future experiments (see our discussion of the related literature below). And though we should not expect a substantial treatment effect in our experiment, if hidden costs of control decrease online then work motivation is also less likely to be reduced by managerial control in WFH arrangements. Another concern of our manipulation is that online data might be of limited quality due to a lower degree of scrutiny compared to the laboratory (Anderhub, Müller, and Schmidt, 2001). Recent experimental studies conclude that online-based and lab-based inferences are equally reliable provided that the same subject pool and similar procedures are used in both settings (Hergueux and Jacquemet, 2015; Normann, Requate, and Waichman, 2014). As detailed in Appendices A and B of the supplementary material, we invested much effort to address this concern and relied on the same subject pool and similar procedures in both treatments. We compare the quality of our online and laboratory data in Appendix F of the supplementary material and we conclude that data of similar quality were collected in the two treatments.

To better mimic naturally-occurring agency relationships, we extend the basic one-shot principal-agent game implemented by Falk and Kosfeld (2006) (henceforth FK) in two major ways. First, we implement a repetitive trial environment where subjects are informed about the payoff consequences of their choices after each repetition of the game which allows them to learn through personal experience and reflection. However, to eliminate the potential confound of repeated-game effects, we match our subjects according to the "no-contagion" protocol in both treatments (Kamecke, 1997). Second, rather than fixing the control level exogenously, we allow the principal to impose either a low or a medium effort level before the agent chooses an effort. This extension of the principal-agent game better reflects the exertion of managerial control in the field and it also enables us to distinguish between the categorical effect of control and the marginal effect of variations in control. Another innovative feature of our design is the incentivized elicitation of subjects' beliefs about the average behavior of their counterpart which is especially helpful to infer principals' motives for their choices.[4]

Our main experimental results are as follows. First, benefits of control outweigh hidden costs in both treatments. Though FK find the opposite result in their original study, most of the experimental

[4]In the principal-agent game the agent's effort corresponds to a monetary cost rather than a real effort. There is however supporting evidence that monetary costs capture well the mental or physical costs in a real effort task (Dutcher, Saral, and Salmon, 2015; Riener and Wiederhold, 2016).

evidence on the magnitude of hidden costs of control confirms our first observation (e.g. Bartling, Fehr, and Schmidt, 2012; Ziegelmeyer, Schmelz, and Ploner, 2012). Second, differences in agents' effort due to an increase in the level of control are significantly larger in the Internet than in the laboratory treatment and these differences originate from higher intrinsic motivation and stronger control aversion in the laboratory. Our second observation suggests that hidden costs of control are reduced in WFH arrangements and it confirms Frey's hypothesis even when the distance in agency relationships is weakly manipulated. Third, we find that effort differences are fairly stable over rounds in both treatments which indicates that even experienced agents react more negatively to the implementation of control in the laboratory. Fourth, both online and laboratory principals correctly expect their agents to be intrinsically motivated and they also correctly expect efforts to increase with the level of control. Finally, principals choose the highest control level most of the time in both treatments which maximizes their monetary payoffs.

Related literature

Our experimental study relates to two strands of the literature.

First, a small experimental literature investigates the impact of workplace arrangements on individual productivity. Dutcher (2012) evaluates the WFH environmental effects on productivity in creative and dull individual tasks. Like in our experiment, half the subjects completed the tasks in the laboratory and the other half outside the laboratory. The results indicate that completing the tasks outside the laboratory affects positively the productivity of creative tasks but negatively the productivity of dull tasks. Bloom, Liang, Roberts, and Ying (2015) conduct a field experiment with call center employees in a Chinese company. Half of the employees invited to the study volunteered to WFH and the other half kept working at the office. WFH lead to a significant increase in the employees' performance. We are not aware of a former study that investigates the impact of workplace arrangements on the hidden costs of control and our experimental study fills this gap.

Second, three experimental studies analyze the influence of the nature of the agency relationship on agents' work motivation. Masella, Meier, and Zahn (2014) test the impact of group identity on the effectiveness of managerial control. Subjects are first assigned to different, artificially created, groups. The authors induce group identity using subjects' painting preferences and they enhance it by having subjects participate in, among others, a quiz where groups compete while group members communicate. Subjects then interact in a principal-agent game (almost) identical to FK's. The results show that hidden costs of control are as strong in between-group agency relationships as in within-group agency relationships though the mechanisms for how control is perceived are group-specific. Riener and Wiederhold (2016) also test the effect of group norms on the hidden costs of control in FK's principal-agent game (except that the principal's choice set is extended to three control levels). They compare a group-building treatment where subjects initially play a coordination game to gain common experience (CE) with an autarky treatment where subjects complete a task in isolation (NCE). They observe higher hidden costs of control in the CE treatment than in the NCE treatment. The fact that their group induction task facilitates gaining positive experience among group members and mutual judgment about this group experience might explain why, contrary to Masella, Meier, and Zahn (2014), their manipulation impacts agents' reactions to control. Dickinson and Villeval (2008) contrast impersonal agency relationships where subjects are matched as strangers and the anonymity of principal-agent pairs is preserved with interpersonal agency relationships where subjects are matched as partners and the pairs anonymity is removed (each pair is allowed to engage in five-minutes of face-to-face social interaction). The agent is engaged in a real-effort task which is likely to generate substantial intrinsic motivation and the principal chooses the probability with which the agent's output is audited. The authors find that the disciplining effect of monitoring dominates the crowding-out effect in both agency relationships and that tighter monitoring by the principal crowds out the agent's effort only in interpersonal agency relationships. Our experiment complements this second strand of the literature which considers a different manipulation of the distance in agency relationships. Future economic experiments should combine our manipulation WFH versus working at the office with induced group identity or interpersonal relationships to increase our understanding of the impact of workplace arrangements on the hidden costs of control.

The next section presents the principal-agent interaction that constitutes the basis of our experimental design and it sheds light on the motivation crowding effects of managerial control. Section 3 outlines our experimental design and procedures and it provides detailed research hypotheses. Section 4 reports our results and Section 5 concludes. The online supplementary material contains six appendices with, among others, the experimental instructions and complementary statistical analyses.

2 Motivation crowding effects of control

We first present the principal-agent interaction that forms the basis of our experimental design. This first subsection details the players' choice sets and the monetary payoffs that result from the interaction. Second, we introduce the players' preferences to derive behavioral predictions for the principal-agent game.

2.1 A principal-agent interaction of managerial control

Consider an agent who engages in a productive activity which is costly to her but beneficial to the principal. Before the agent exerts effort, the principal can either decide to leave the agent's effort set unrestricted by choosing "no control" ($\underline{e} = 1$) or he can decide to restrict the agent's effort set by choosing one of two control levels: "low control" ($\underline{e} = 2$) or "medium control" ($\underline{e} = 3$). The agent then chooses an effort level $e \in \{\underline{e}, \underline{e} + 1, \ldots, 10\}$. Table 1 shows the monetary payoffs (in experimental currency units) where the fair and most efficient effort level locates slightly above the middle ($e = 7$).

	Effort level									
	1	2	3	4	5	6	7	8	9	10
Agent's monetary payoffs	99	98	96	93	89	83	75	65	51	35
Principal's monetary payoffs	1	16	29	41	53	64	75	82	87	90

Table 1: Monetary payoffs by effort level.

Several considerations led to the players' monetary payoffs shown in Table 1. First, exerting more effort than the minimal one is cheap for the agent and extremely beneficial for the principal. This feature of the monetary payoffs ensures that even agents with little intrinsic motivation can express their willingness to perform in the interest of the principal when managerial control is absent ($\underline{e} = 1$). Consequently, hidden costs of control are likely to prevail in our experiment (i.e., the crowding out of intrinsic motivation when $\underline{e} > 1$). Second, effort costs are assumed to be convex since exerting low effort at work is usually not very costly but once the agent is working to capacity marginal effort costs become tremendous. Third, benefits from the agent's effort are assumed to be concave which reflects productivity losses due to physical restrictions.

2.2 Behavioral predictions

The aim of this subsection is to shed light on the motivation crowding effects of managerial control. We therefore focus on the agent's behavior and endow her with flexible preferences. Concretely, we allow for the possibility that the agent has an intrinsic motivation to perform in the interest of the principal and that control may either substitute for (crowding out) or complement (crowding in) the agent's intrinsic motivation. On the other hand, we simply assume that the principal maximizes his own monetary payoffs and that he is able to perfectly forecast the agent's effort at any control level.

We adopt the specification of *control-state-preferences* due to Bowles and Polanía-Reyes (2012) where preferences are state-dependent with different control levels constituting different states.[5] For a given couple $(\underline{e}, e) \in \{1,2,3\} \times \{\underline{e}, \ldots, 10\}$, the agent's utility is

$$u_A\left(e, \underline{e}; \lambda_0, \lambda_c, \lambda_m\right) = \Pi_A(e) + e\,\lambda_0\left(1 + \lambda_c\,\mathbb{I}\{\underline{e} > 1\} + \lambda_m\left(\underline{e} - 1\right)\right), \tag{1}$$

where $\Pi_A(\cdot)$ is the agent's monetary payoff function, $\lambda_0 \geq 0$ measures the agent's intrinsic motivation, $\lambda_c \in \mathbb{R}$ measures the categorical effect of control on motivation, $\lambda_m \in \mathbb{R}$ measures the marginal effect of variations in control on motivation, and the indicator $\mathbb{I}\{\underline{e} > 1\}$ equals 1 if $\underline{e} > 1$ and zero otherwise. The agent's intrinsic motivation can either reflect other-regarding preferences that induce her to help the principal more than an own-monetary payoff maximizing agent would, intrinsic preferences for autonomy in decision-making, the desire to obtain social esteem, etc.

The control-state-preferences specification encompasses the standard case where variations in the agent's intrinsic motivation affect her chosen effort independently of the presence or magnitude of control ($\lambda_c = \lambda_m = 0$). The specification also allows intrinsic motivation to be either heightened by the presence ($\lambda_c > 0$) or magnitude ($\lambda_m > 0$) of control or, the more commonly considered case, affected adversely by the presence ($\lambda_c < 0$) or magnitude ($\lambda_m < 0$) of control. For the sake of brevity, we do not make explicit the foundations for the agent's control-state-preferences. In the theoretical literature on hidden costs of control, crowding effects arise either because the agent infers from the enforced level of control the motivations of the principal or his beliefs regarding the agent's motivation, or because a restricted set of efforts undermines the agent's sense of autonomy.[6] Finally, we assume that whenever the agent is indifferent between several effort levels she chooses the least costly one, and that whenever the principal is indifferent between several control levels he chooses the most restrictive one.

For a given control level $\underline{e} \in \{1,2,3\}$, the agent's best reply consists in exerting effort

- $e^*(\underline{e}) = \underline{e}$, the lowest effort, if $\lambda_0\left(1 + \lambda_c\,\mathbb{I}\{\underline{e} > 1\} + \lambda_m\left(\underline{e} - 1\right)\right) \leq \underline{e}$;
- $e^*(\underline{e}) \in \{\underline{e}{+}1, \ldots, 9\}$, a "mid" effort, if $\lambda_0\left(1 + \lambda_c\,\mathbb{I}\{\underline{e} > 1\} + \lambda_m\left(\underline{e} - 1\right)\right) \in]\Pi_A(e^*(\underline{e}) - 1) - \Pi_A(e^*(\underline{e})),$ $\Pi_A(e^*(\underline{e})) - \Pi_A(e^*(\underline{e}) + 1)]$; and
- $e^*(\underline{e}) = 10$, the highest effort, if $\lambda_0\left(1 + \lambda_c\,\mathbb{I}\{\underline{e} > 1\} + \lambda_m\left(\underline{e} - 1\right)\right) > 16$.

Table 2 illustrates the agent's best reply where intrinsic motivation is either weak, moderate or strong, and various crowding effects are considered.

[5]Bowles and Polanía-Reyes (2012) survey fifty experiments where (explicit economic) incentives and social preferences are either substitutes (incentives crowd out social preferences) or complements (incentives crowd in social preferences). Accordingly, they introduce a model of *incentive-state-preferences* where preferences are state-dependent with different incentive levels constituting different states. Control of effort is one form of explicit economic incentives that principals use to discipline agents.

[6]Information-based foundations for control-state-preferences in agency relationships are provided in, among others, Sliwka (2007) and von Siemens (2013).

	Intrinsic motivation								
	Weak ($\lambda_0 \leq 1$)			Moderate ($\lambda_0 = 3.5$)			Strong ($\lambda_0 = 9.5$)		
	$\underline{e} = 1$	$\underline{e} = 2$	$\underline{e} = 3$	$\underline{e} = 1$	$\underline{e} = 2$	$\underline{e} = 3$	$\underline{e} = 1$	$\underline{e} = 2$	$\underline{e} = 3$
No crowding effects ($\lambda_c = \lambda_m = 0$)	1	2	**3**	4	4	**4**	7	7	**7**
Low categorical & high marginal crowding in ($\lambda_c = 0.075, \lambda_m = 0.425$)	1	2	**3**	4	5	**6**	7	9	**10**
High categorical & low marginal crowding in ($\lambda_c = 0.425, \lambda_m = 0.075$)	1	2	**3**	4	5	**5**	7	9	**9**
Low categorical crowding out & high marginal crowding in ($\lambda_c = -0.075, \lambda_m = 0.425$)	1	2	**3**	4	5	**6**	7	8	**10**
Low categorical & low marginal crowding out ($\lambda_c = -0.075, \lambda_m = -0.075$)	1	2	**3**	**4**	3	3	7	**7**	6
High categorical crowding out & low marginal crowding in ($\lambda_c = -0.425, \lambda_m = 0.075$)	1	2	**3**	**4**	3	3	**7**	6	6
Low categorical & high marginal crowding out ($\lambda_c = -0.075, \lambda_m = -0.425$)	1	2	**3**	**4**	2	3	**7**	5	3
High categorical & low marginal crowding out ($\lambda_c = -0.425, \lambda_m = -0.075$)	1	2	**3**	**4**	2	3	**7**	5	5

Notes: Bold numbers highlight principal's best reply.

Table 2: An illustration of the agent's best reply.

As he is able to perfectly forecast the agent's effort, the principal enforces the medium control level in the following three cases: i) Intrinsic motivation is sufficiently weak which leads the agent to exert minimal effort for each level of control enforced by the principal; ii) Intrinsic motivation is sufficiently strong and crowding effects are sufficiently low which leads the agent to exert identical effort whatever the level of control enforced by the principal; and iii) $\lambda_c + 2\lambda_m$ is not too negative which leads the agent to exert no higher effort than under medium control. The principal enforces the low control level if the agent's intrinsic motivation is sufficiently strong and $\lambda_c + \lambda_m$ is not too negative but λ_m is sufficiently negative. Finally, the principal refrains from restricting the agent's effort set if her intrinsic motivation is sufficiently strong and both $\lambda_c + \lambda_m$ and $\lambda_c + 2\lambda_m$ are sufficiently negative.

3 Experimental design, procedures, and hypotheses

3.1 Design

In both treatments subjects repeatedly take part in the principal-agent interaction described in Section 2.1 where the payoffs in Table 1 are in experimental currency units. We employ the strategy method, meaning that the agent makes her choice for each of the three control levels before knowing the principal's actual decision.[7] Concretely, each agent is asked to choose a triplet of effort levels $(e(1), e(2), e(3))$ where $e(1) \in \{1, 2, \ldots, 10\}$ is payoff-relevant in case the principal does not enforce a minimal effort, $e(2) \in \{2, 3, \ldots, 10\}$ is payoff-relevant in case the principal enforces a low effort, and $e(3) \in \{3, 4, \ldots, 10\}$ is payoff-relevant in case the principal enforces a medium effort.

[7]To avoid demand effects, we did not distinguish in the instructions and on the decision screens between "no control" and a level of control which restricts the agent's choice set. All three control levels were phrased in the same way, namely that the principal forces the agent to exert an effort level of at least 1, 2 or 3. Thus, our instructions are conservative with respect to the categorical effect of control. See the experimental screens in Appendix B.3.

In a given session, each subject is assigned either the role of agent or the role of principal. Subjects gain experience with the context and the behavior of others during 10 repetitions of the interaction. Roles are kept constant over all rounds. The matching follows a "no-contagion" protocol which suppresses repeated-game effects.[8] Indeed, other-regarding considerations are confounded with strategic ones in the presence of repeated-game effects.

Belief elicitation

Before they interact in the agency relationship, subjects are asked to guess the average behavior of their counterpart. In each round, subjects make three guesses. Principals are asked to guess, for each control level, the average effort that will be chosen by agents (since we employ the strategy method, for each control level *all* agents choose an effort). Each principal reports his guesses by keying in a vector $(b_P(1), b_P(2), b_P(3))$ with $\underline{e} \leq b_P(\underline{e}) \leq 10$. Agents are asked to guess, for each control level, the natural frequency of principals that will chose the respective control level. Each agent reports her guesses by keying in a vector $(b_A(1), b_A(2), b_A(3))$ with $0 \leq b_A(\underline{e}) \leq 100$ and $b_A(1) + b_A(2) + b_A(3) = 100$.

We limit the possibility to learn about the choices of other subjects. Once all guesses and choices have been made in a given round, subjects are only informed about the behavior of their counterpart. Subjects do not learn about the correctness of their guesses during the experimental session.

Earnings

Each subject is paid a flat amount of 30 experimental currency units (ECUs) for completing a survey. As far as the interaction part is concerned, only one of the 10 rounds is payoff-relevant. The payoff-relevant round is randomly selected at the end of the session. For each subject, depending on the outcomes of random draws, either one of the three guesses is paid or the earnings are determined by the agency relationship. The randomly chosen guess is paid according to the following scheme: If an agent's (principal's) guess differs by no more than 5 percentage points (0.5 effort levels) from the true value then the subject earns 70 ECUs. Otherwise the subject earns 20 ECUs.

3.2 Practical procedures

Both treatments were conducted with the help of an Internet platform developed by the authors and detailed in Appendix A. All 440 subjects were students who had agreed to participate in economic experiments. The data were collected in two waves. Two sessions per treatment were first conducted in Jena with 106 students from the University of Jena (November 2010 and January 2011). We then conducted another six sessions per treatment in Konstanz with 334 students from the University of Konstanz (November 2014 and April 2015). All sessions followed exactly the same procedures.[9]

Subjects were invited using the ORSEE recruitment system (Greiner, 2015). Students received an invitation email with a link to a registration page. On this page they were informed about the general rules of the study, and about the fact that the other participants are those they usually interact with in the laboratory. For registration students had to enter some information (gender, month and year of birth, nationality, mother tongue, and email address). Each student could register only once. Registered

[8]No subject i is ever matched with a subject who has previously been matched with someone who has been matched with someone subject i has already been matched with (and for any positive integer n the sentence which replaces the phrase "who has previously been matched with someone who has been matched with someone" in the previous sentence with n copies of the same phrase is also true).

[9]The first laboratory sessions took place in the experimental laboratory of the Max Planck Institute of Economics in Jena while the next laboratory sessions took place in *Lakelab*, the laboratory for economic experiments in Konstanz. Both laboratories strictly adhere to a non-deception policy.

subjects received a survey token via email. Answering the survey questions took on average 10 minutes and subjects had a time frame of a few days to do so. In both treatments subjects completed the survey at their place of choice (e.g. at home).[10]

Subjects who completed the survey could register for an experimental session and received a session token to the experiment via email. To circumvent a potential impact of the survey on choices in the interactive part experimental sessions were conducted on a later day. Each session took slightly more than one hour.

In the online treatment there was a prearranged start time for each of the eight sessions which took place in the afternoon or evening, and each of the 232 subjects had to log on not later than that time. Like for the survey, subjects made their interactive choices at their place of choice. The eight sessions of the laboratory treatment took place in the afternoon or evening with a total of 208 subjects. Instructions were not read aloud.

In both treatments subjects received their earnings in the laboratory. In the online treatment subjects were informed that they would receive a compensation fee for collecting their earnings which corresponds to the usual show-up fee in the two locations (2.50 euros in Jena and 3 euros in Konstanz). In the laboratory treatment we also added the usual show-up fee to subjects' earnings.[11] In Jena, 1 ECU was converted to 0.15 euros. To adjust for differences in purchasing power, absolute earnings were slightly raised in Konstanz where 1 ECU was converted to 0.20 euros. Subjects in Jena (Konstanz) earned 15.42 (19.60) euros on average (about 21 (22) US dollars at the time of the sessions).[12] The participation process is detailed in Appendix A and screens with the instructions are provided in Appendix B.

Unsurprisingly, the socio-demographic characteristics of subjects are very similar in the two treatments. 52% and 49% of the subjects are female and mean age is 21.5 and 22.1 years in the laboratory and online treatment respectively. In the laboratory (online) treatment, the sample composition according to the field of study is such that 31% (34%) of the subjects belong to the category "business administration & economics", 28% (31%) belong to the category "other behavioral & social sciences", 19% (18%) belong to the category "humanities" and 23% (18%) belong to the category "engineering, life & natural sciences".[13]

3.3 Hypotheses

Our first and main research hypothesis is (a weak version) of Frey's hypothesis.

Hypothesis 1 [effort differences]. Differences in agents' effort due to an increase in the level of control are larger online than in the laboratory.

Concretely, we postulate that the difference between the agent's effort under low control ($e(2)$) and the agent's effort under no control ($e(1)$) as well as the difference between the agent's effort under medium

[10]We do not elaborate on the survey as it is not central to the study at hand.

[11]Subjects in the laboratory were paid immediately after the session as usual while online subjects collected their earnings about three days later, once the official lottery draw had determined their payoffs. Accordingly, monetary incentives are less salient in the online than in the laboratory treatment which might add to the weakness of our manipulation.

[12]There is still a difference in purchasing power between East and West Germany. The purchasing power index of Thuringia (federal state of Jena, East Germany) is about 84 whereas the purchasing power index of Baden-Württemberg (federal state of Konstanz, West Germany) amounts to 107 (GfK, 2010, 2014). The different payoffs in euros in the two locations hardly translate into different payoffs in US dollars because of fluctuations in the conversion rates between the two currencies.

[13]In one laboratory session in Jena, two participants dropped out because of a network connection failure. In the online treatment, three subjects in Konstanz dropped out. For each dropout, one participant of the other role was excluded from further participation. Participants who dropped out or were excluded were informed privately on their screens while the session proceeded smoothly for the remaining participants. We restrict our data analysis to participants who completed the experiment.

control ($e(3)$) and the agent's effort under low control ($e(2)$) are strictly larger in the Internet than in the laboratory treatment:

$$\text{a) } e(2) - e(1)\big|_{Internet} > e(2) - e(1)\big|_{Laboratory} \quad \text{; and}$$

$$\text{b) } e(3) - e(2)\big|_{Internet} > e(3) - e(2)\big|_{Laboratory} \quad .$$

We therefore expect that hidden costs of control are reduced online (assuming of course that they exist in both treatments). Note that our first hypothesis is consistent with agents expressing non-negligible intrinsic motivation in the two treatments and stronger control aversion in the laboratory than online, but also with the absence of control aversion in each treatment and non-negligible intrinsic motivation only in the laboratory. Fortunately, we can distinguish between intrinsic motivation and control aversion in our data analysis as we rely on the strategy method to elicit the agents' chosen efforts. Thus, we decompose Hypothesis 1 into its intrinsic motivation and control aversion components.

Hypothesis 1a [intrinsic motivation]. In the absence of control, agents exert less effort online than in the laboratory: $e(1)\big|_{Internet} < e(1)\big|_{Laboratory}$.

Hypothesis 1b [control aversion]. Control aversion is weaker online than in the laboratory.

A control averse agent chooses a higher effort in the absence of control than if controlled. Concretely, aversion to low control is expressed by $e(1) > e(2)$ and aversion to medium control is expressed by $e(2) > e(3)$. To isolate the effects of control aversion from disciplining effects of control, we have to rule out opportunistic choices in the effort distributions (as rightly argued by FK). Any effort smaller than the minimal effort level $e(1) < \underline{e} = 2$ and $e(2) < \underline{e} = 3$ is shifted to the minimal effort level $\underline{e}$. This means that opportunistic effort choices under no control ($e(1) = 1$) are set equal to 2, and opportunistic choices under low control ($e(2) = 2$) are set equal to 3. Aversion to low control implies that the difference between efforts under low control and the shifted efforts under no control ($e(2) - max[e(1), 2]$) is negative and aversion to medium control implies that the difference between efforts under medium control and the shifted efforts under low control ($e(3) - max[e(2), 3]$) is negative. Thus, Hypothesis 1b states that:

$$\text{a) } e(2) - max[e(1), 2]\big|_{Internet} > e(2) - max[e(1), 2]\big|_{Laboratory} < 0 \quad \text{; and}$$

$$\text{b) } e(3) - max[e(2), 3]\big|_{Internet} > e(3) - max[e(2), 3]\big|_{Laboratory} < 0 \quad .$$

FK observe that a majority of their principals chooses not to control the agent which suggests that they anticipate the existence of hidden costs of control. In line with this finding and assuming that our first hypothesis is confirmed, we conjecture that agency relationships are characterized by lower effort discretion online than in the laboratory.

Hypothesis 2 [control level]. Principals enforce larger control levels online than in the laboratory: $\underline{e}\big|_{Internet} > \underline{e}\big|_{Laboratory}$.

4 Results

We first compare behavior in Jena and Konstanz and, given the absence of significant differences, we base our further analysis on the pooled data of the two locations. Next, we compare intrinsic motivation and crowding effects at the aggregate level in the two treatments. Finally, we formally test our two research hypotheses.

4.1 Comparing subjects' behavior in Jena and Konstanz

Appendix C provides detailed comparisons of the two locations with respect to the choices and beliefs of subjects. The distribution of p-values derived from 42 tests is very close to the uniform distribution. We conclude that behavior in the two locations does not differ in any meaningful way. Further statistical analysis is based on the pooled data of the two locations.[14]

4.2 Intrinsic motivation and crowding effects

Table 3 shows agents' efforts as a function of the control level in the two treatments. In both panels the first row reports the average effort and the second row reports standard deviation followed by 1st quartile followed by median followed by 3rd quartile for each control level. We observe that the average effort in the absence of control is around the largest minimum effort enforceable by the principal, which indicates the presence of sizable intrinsic motivation, and that the average effort increases with the control level in each treatment.[15] Interestingly, increases in average effort are larger online than in the laboratory.

	No control	Low control	Medium control
Laboratory	**3.46**	**3.51**	**3.63**
(1040 observations)	(2.73;1;2;7)	(1.87;2;2;5)	(1.21;3;3;4)
Internet	**2.79**	**3.15**	**3.60**
(1160 observations)	(2.30;1;1;5)	(1.61;2;2;4)	(1.20;3;3;4)

Table 3: Agents' efforts as a function of the control level.

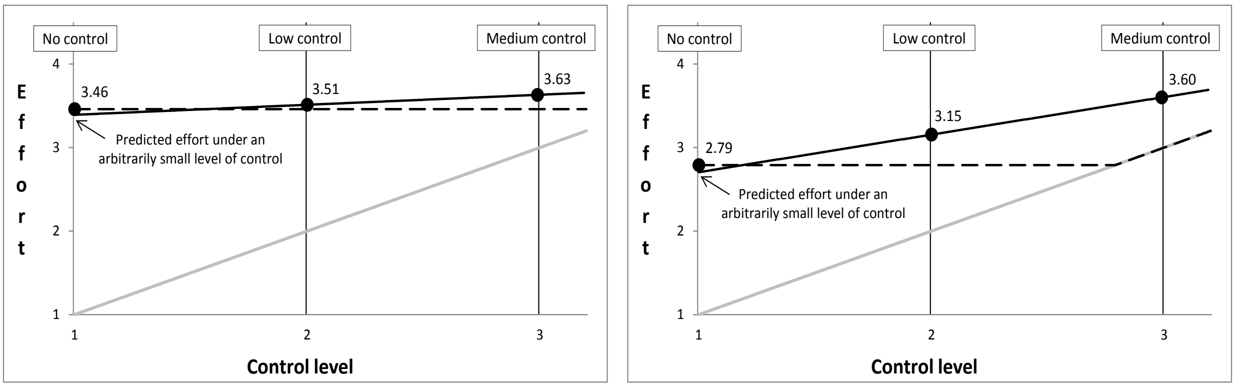
To infer the extent of intrinsic motivation and crowding effects at the aggregate level in each treatment, we here assume that average efforts are best replies of a representative agent endowed with control-state-preferences (details are provided in Appendix D). First, we infer from average efforts in the absence of control that intrinsic motivation is moderate in each treatment and weaker online than in the laboratory ($\hat{\lambda}_0|_{Internet} = 2.69 < \hat{\lambda}_0|_{Laboratory} = 3.69$). Second, increases in average efforts due to a variation in the control level from low to medium control (0.45 and 0.12 in the Internet and laboratory treatment respectively) enable us to infer that $\hat{\lambda}_m|_{Internet} = 0.25 > \hat{\lambda}_m|_{Laboratory} = 0.05$. Accordingly, there is *marginal crowding in* in each treatment which is much larger online than in the laboratory. Third, given average efforts under low and medium control as well as the already inferred values of λ_0 and λ_m, the optimal effort of the representative agent following an arbitrarily small level of control is 2.70 and 3.39

[14] In a nutshell, we find that: i) in both treatments, the null hypothesis that differences in agents' effort between control levels are the same in Jena and Konstanz is never rejected at the 10 percent level whether the first half, second half or the two halves of the session are considered; ii) in the laboratory treatment, we do not reject the null hypothesis that the distribution of agents' beliefs across all rounds is the same in Jena and Konstanz at the 5 percent level; iii) in the Internet treatment, we do not reject the null hypothesis that the distribution of agents' beliefs across all rounds is the same in Jena and Konstanz at the 10 percent level; iv) in both treatments, the null hypothesis that control levels chosen by principals are the same in Jena and Konstanz is never rejected at the 10 percent level whether the first half, second half or the two halves of the session are considered; v) in the laboratory treatment, we do not reject the null hypothesis that the distribution of principals' beliefs across all rounds is the same in Jena and Konstanz at the 5 percent level; and vi) in the Internet treatment, we do not reject the null hypothesis that the distribution of principals' beliefs across all rounds is the same in Jena and Konstanz at the 10 percent level.

[15] If the principal cares solely about his monetary payoffs then he should limit the agent's effort discretion to the largest possible extent. The fact that the variability in efforts reduces as the level of control increases provides another rationale for principals to select the medium control level.

in the Internet and laboratory treatment respectively. We therefore infer that $\hat{\lambda}_c\big|_{Internet} = -0.05 < \hat{\lambda}_c\big|_{Laboratory} = -0.03$ meaning that managerial control *categorically crowds out* 5 and 3 percent of the effect of intrinsic motivation online and in the laboratory respectively.

As a summary, Figure 1 shows the observed average effort and illustrates the predicted effort for the representative agent as a function of the control level in each treatment. Grey lines show the predicted efforts for a purely self-interested representative agent ($\lambda_0 = 0$), black dashed lines show the predicted efforts for a representative agent who is intrinsically motivated ($\hat{\lambda}_0\big|_{Internet} = 2.69$ and $\hat{\lambda}_0\big|_{Laboratory} = 3.69$) but indifferent to the presence or magnitude of control ($\lambda_c = \lambda_m = 0$), and black solid lines show the predicted efforts for a representative agent with inferred parameter values.



(a) Laboratory

(b) Internet

Figure 1: Average and predicted efforts as a function of the control level.

4.3 Testing Hypothesis 1: Effort differences in the two treatments

To formally test whether differences in agents' effort due to an increase in the level of control are larger online than in the laboratory, we rely on a series of regression models. The estimation method is linear mixed models where random intercepts at the agent and session levels are included. Random effects are assumed to be independent and to follow a normal distribution with mean zero. Our regression models allow the agent's behavior to be correlated across rounds as well as the behavior of different agents in the same session to be correlated.[16] Models 1 to 3 predict effort differences under low and no control while models 4 to 6 predict effort differences under medium and low control. Table 4 summarizes the estimation results.

In models 1 and 4, the effort difference is regressed against an intercept and the experimental condition where the dummy variable *Int* identifies the Internet treatment. Averaged over all agents and rounds, the effort difference is always positive but not significantly different from zero in the laboratory treatment and the difference always increases significantly in the Internet treatment. Regression results therefore show that differences in agents' effort due to an increase in the level of control are significantly larger online than in the laboratory.

The dynamics of effort differences indicate that the behavior of experienced agents is also supportive of our first hypothesis. Models 2 and 5 extend models 1 and 4 by including a dummy *Half2* for the second half of rounds (rounds 6-10), its interaction with the experimental condition *Int*Half2*, and a difference in beliefs ($b_A(2)-b_A(1)$ and $b_A(3)-b_A(2)$ respectively). For both effort differences, the estimated coefficients

[16]For all regression models, the estimate of the random intercept at the session level is basically zero. The nature of our matching protocol limits the degree to which the behavior of agents in the same session is correlated. Running the same regressions without a random intercept at the session level generates almost identical standard errors.

	Dependent variable: Difference between effort under					
	low control and no control $e(2)-e(1)$			medium and low control $e(3)-e(2)$		
Model	(1)	(2)	(3)	(4)	(5)	(6)
$Constant$	0.050	0.088	0.249	0.114	0.063	0.130
	(0.101)	(0.106)	(0.173)	(0.094)	(0.102)	(0.162)
Int	0.309**	0.286**	0.261*	0.332**	0.318**	0.298**
	(0.139)	(0.146)	(0.145)	(0.129)	(0.132)	(0.132)
$Half2$		-0.062	-0.062		-0.045	-0.045
		(0.061)	(0.061)		(0.051)	(0.051)
$Int * Half2$		0.047	0.047		0.046	0.047
		(0.084)	(0.084)		(0.069)	(0.069)
$b_A(2)-b_A(1)$		-0.002	-0.002			
		(0.002)	(0.002)			
$b_A(3)-b_A(2)$					0.001*	0.001*
					(0.001)	(0.001)
Age			-0.135			-0.039
			(0.139)			(0.128)
$Male$			-0.025			0.084
			(0.141)			(0.129)
$Social$			-0.092			0.015
			(0.178)			(0.164)
Hum			0.289			-0.003
			(0.226)			(0.207)
$Tech$			-0.330*			-0.336**
			(0.183)			(0.168)
Observations	2,200	2,200	2,200	2,200	2,200	2,200
Log-likelihood	-3349.874	-3348.781	-3344.465	-2939.273	-2937.337	-2934.404

Notes: Standard errors in parentheses. ***(1%); **(5%); *(10%) significance level.

Table 4: Determinants of effort differences.

of *Int* are significantly positive while those of *Half2* and *Int*Half2* do not significantly differ from zero. Thus, effort differences and the influence of the distance in agency relationships are fairly stable over time.[17] Estimated coefficients of belief differences are either non-significantly different from zero or small which is hardly surprising since agents can condition their effort on the control level. Figure 2 displays the effort differences over time.

The estimated coefficients of models 2 and 5 barely change when demographic controls are included (age, gender and academic major)[18] as shown in columns (3) and (6) of Table 4. Moreover none of the demographics significantly impacts effort differences except for one of the field of study.[19]

[17] Linear combinations of the coefficients Int and $Int * Half2$ are significantly different from zero, confirming that in the second half, effort differences are also larger online than in the laboratory (χ^2 tests: $p-values < 0.05$).

[18] *Age* equals 1 if the subject's age is at the median or above and 0 otherwise. *Social* equals 1 if the field of study belongs to the category "behavioral & social sciences except economics" and 0 otherwise. *Hum* equals 1 if the field of study belongs to "humanities" and 0 otherwise. *Tech* equals 1 if the field of study belongs to "engineering, life & natural sciences" and 0 otherwise.

[19] The significant impact of $Tech$ on effort differences is not robust when interacting the field of study with the experimental condition. Additional regression results are reported in Appendix E.2. Note that our demographic controls affect the impact of the experimental condition on the difference between effort under low and no control for economists (constant) in the first half of rounds where the estimated coefficient of Int is only weakly significant in model 3 ($p-value = 0.070$). Linear combinations of the coefficients Int and $Int * Half2$ differ from zero at the 5 percent level (χ^2 test: $p-value = 0.033$)

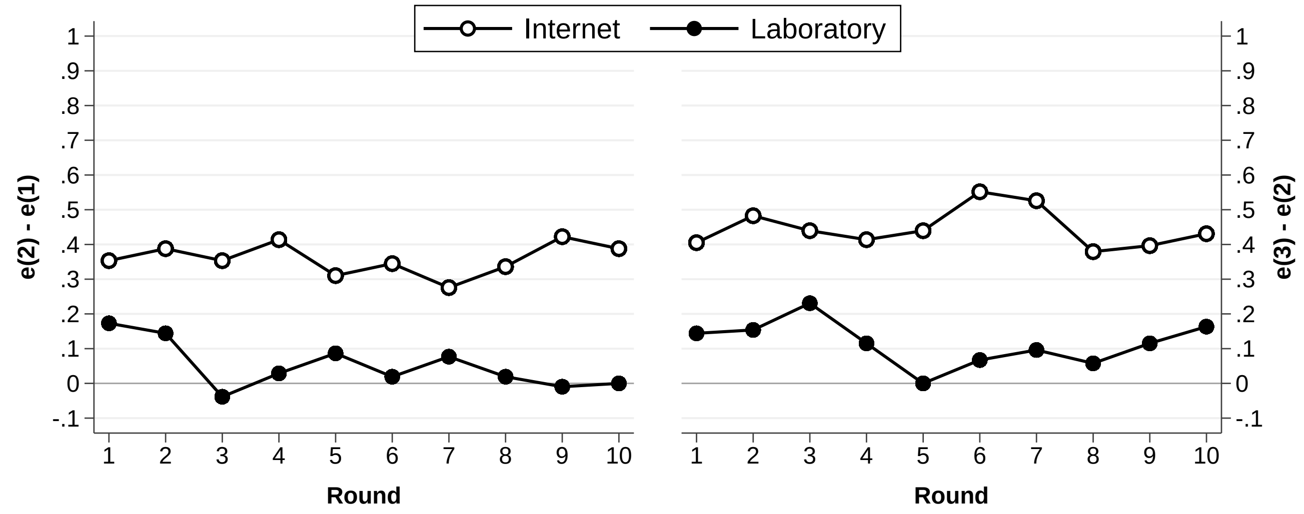


Figure 2: Effort differences over time.

In sum, effort differences are positive in each treatment meaning that disciplining effects outweigh crowding effects of managerial control. Most importantly, the evidence supports our first hypothesis since effort differences are significantly larger online than in the laboratory. We now shed light on the mechanisms driving this finding.[20]

Testing Hypothesis 1a: Intrinsic motivation in the two treatments

We formally test whether in the absence of control agents exert less effort online than in the laboratory by relying on a series of regression models. As before, our regression models are linear mixed models with random intercepts for agents and sessions and Table 5 reports the estimation results.

In model 1 agents' effort under no control is regressed against an intercept and the experimental condition. Averaged over all rounds, the "no control" effort exceeds the largest effort level enforceable by the principal in the laboratory and it is significantly lower in the Internet treatment. Still, the average effort of online agents is significantly greater than the minimal effort level of 1 (χ^2 test: $p-value < 0.01$). Model 2 shows that the difference in the agent's intrinsic motivation between the two treatments is fairly stable over time (the coefficients of $Half2$ and $Int * Half2$ do not significantly differ from zero) and that the impact of beliefs is either insignificant or small. The estimated coefficients of model 2 are hardly affected when including demographic controls as shown in column (3) of Table 5. Except for one of the field of study none of the demographics significantly impacts the agent's effort under no control.[21]

Our regression results on the agent's effort in the absence of control reveal that intrinsic motivation is present and rather stable over time in each treatment, and that it is significantly weaker online than in the laboratory which supports Hypothesis 1a.

meaning that the effort difference $(e(2) - e(1))$ is significantly larger online than in the laboratory in the second half of the session.

[20]Since differences in effort levels are not linear with respect to the agent's effort costs, we performed robustness checks with differences in effort costs as the dependent variable. Appendix E.3 reports the results of these robustness checks which are in line with the regression results contained in the main text.

[21]Our demographic controls affect the impact of the experimental condition on efforts under no control for economists (constant) in the first half of rounds where the estimated coefficient of Int is only weakly significant in model 3 ($p-value = 0.064$). Linear combinations of Int and $Int * Half2$ differ from zero at the 5 percent level (χ^2 test: $p-value = 0.015$) meaning that efforts under no control are significantly smaller online than in the laboratory for the second half of rounds. Robustness checks on agents' effort costs under no control confirm our findings (see Appendix E.4).

	Dependent variable: Effort under no control $e(1)$		
Model	(1)	(2)	(3)
Constant	3.462***	3.201***	2.873***
	(0.214)	(0.327)	(0.434)
Int	-0.668**	-0.606**	-0.541*
	(0.295)	(0.294)	(0.292)
$Half2$		-0.021	-0.021
		(0.083)	(0.083)
$Int * Half2$		-0.173	-0.173
		(0.112)	(0.112)
$b_A(2)$		0.013***	0.013***
		(0.004)	(0.004)
$b_A(3)$		0.001	0.001
		(0.003)	(0.003)
Age			-0.095
			(0.289)
Male			0.095
			(0.292)
Social			0.117
			(0.370)
Hum			0.157
			(0.469)
Tech			0.964**
			(0.380)
Observations	2,200	2,200	2,200
Log-likelihood	-4091.993	-4077.437	-4073.722

Notes: Standard errors in parentheses.
***(1%); **(5%); *(10%) significance level.

Table 5: Agents' effort in the absence of control.

Testing Hypothesis 1b: Control aversion in the two treatments

The regression models on control aversion are identical to those on effort differences except for the dependent variable. Models 1 to 3 predict aversion to low control, which corresponds to the difference between the effort under low control and the shifted effort under no control ($e(2) - max[e(1), 2]$), and models 4 to 6 predict aversion to medium control, which corresponds to the difference between the effort under medium control and the shifted effort under low control ($e(3) - max[e(2), 3]$). Table 6 reports our estimation results on control aversion.

In models 1 and 4 control aversion is regressed against an intercept and the experimental condition. Averaged over all agents and rounds, control aversion is always significantly negative in the laboratory and significantly less so online. Yet, online aversion to low (medium) control is significantly different from zero at the 5 (10) percent level (χ^2 tests: $p - value = 0.018$ and $p - value = 0.085$). Moreover, the results of models 2 and 5 show that experienced agents tend to express somewhat more control aversion than inexperienced agents in both treatments and that belief differences hardly matter. Figure 3 displays the extent of control aversion over time. Finally, the inclusion of demographic controls has little impact on the estimated coefficients of models 2 and 5 and none of the demographics consistently affects control aversion as shown in columns (3) and (6) of Table 6.

	Dependent variable: Aversion to					
	low control $e(2) - max[e(1), 2]$			medium control $e(3) - max[e(2), 3]$		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.435***	-0.379***	-0.247*	-0.413***	-0.359***	-0.319***
	(0.078)	(0.083)	(0.132)	(0.068)	(0.077)	(0.120)
Int	0.261**	0.242**	0.222**	0.303***	0.284***	0.266***
	(0.107)	(0.113)	(0.111)	(0.093)	(0.098)	(0.098)
Half2		-0.106*	-0.106*		-0.103**	-0.103**
		(0.054)	(0.054)		(0.044)	(0.044)
$Int * Half2$		0.038	0.038		0.038	0.038
		(0.074)	(0.074)		(0.060)	(0.060)
$b_A(2) - b_A(1)$		-0.001	-0.001			
		(0.001)	(0.001)			
$b_A(3) - b_A(2)$					-0.000	-0.000
					(0.001)	(0.001)
Age			-0.117			-0.031
			(0.105)			(0.094)
Male			-0.055			0.066
			(0.106)			(0.095)
Social			-0.092			0.001
			(0.135)			(0.120)
Hum			0.331*			0.071
			(0.171)			(0.152)
Tech			-0.220			-0.239*
			(0.139)			(0.123)
Observations	2,200	2,200	2,200	2,200	2,200	2,200
Log-likelihood	-3048.350	-3045.536	-3039.528	-2603.731	-2599.655	-2596.600

Notes: Standard errors in parentheses. ***(1%); **(5%); *(10%) significance level.

Table 6: Determinants of control aversion.

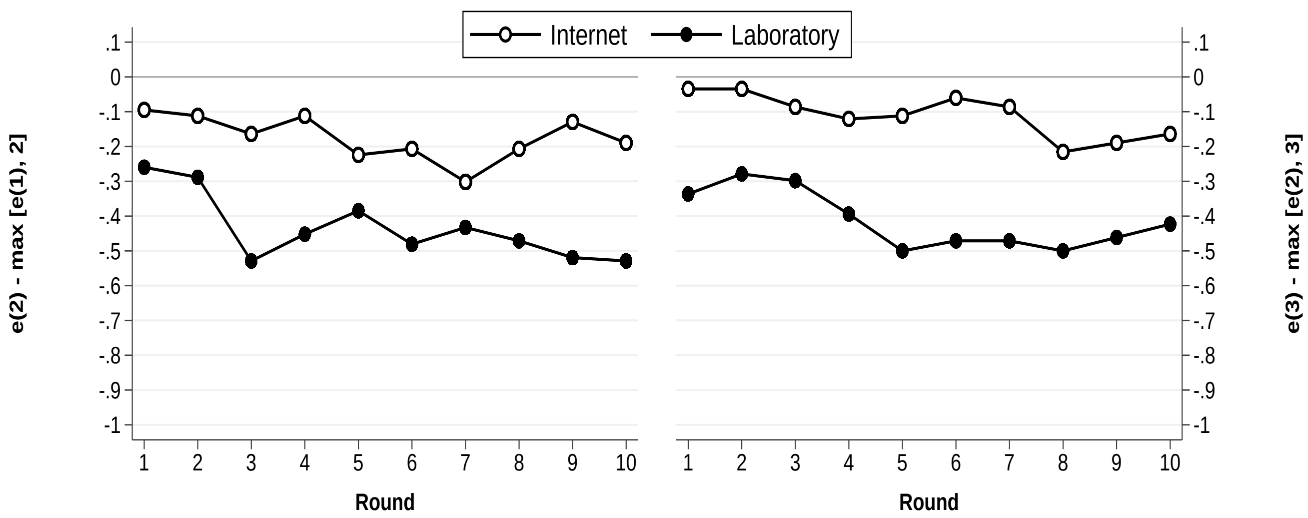


Figure 3: Control aversion over time.

In sum, managerial control significantly crowds out intrinsic motivation in both treatments, negative reactions to control tend to increase over time, and hidden costs of control are lower online than in the laboratory which supports Hypothesis 1b.[22]

Agents' beliefs

Averaged over all agents and rounds, expected frequencies of no control, low and medium control are 14%, 19% and 67% (12%, 17% and 71%) in the Internet (laboratory) treatment. We fail to reject the null hypothesis that the distribution of individual beliefs, averaged across rounds for each agent, about the frequency of principals who choose the no, low or medium control level is the same in the two treatments (Wilcoxon rank-sum test: $p-values > 0.1$). Differences between agents' beliefs in the two treatments are also non-significant in the second half of rounds (Wilcoxon rank-sum test: $p-values > 0.1$) and agents expect slightly higher effort control over time. In the second half of rounds, expected frequencies of no control, low and medium control are 13%, 16% and 71% (11%, 14% and 75%) in the Internet (laboratory) treatment. To assess the correctness of agents' beliefs, we compute mean squared differences between beliefs and actual frequencies of control levels and we average them across rounds for each agent. When considering all or the second half of rounds, we never reject the null hypothesis that the distribution of the mean squared difference is the same in the two treatments (Wilcoxon rank-sum test: $p-values > 0.1$). Experienced agents do not predict principals' control decisions significantly better in any of the two treatments. Further details about agents' beliefs are given in Appendix E.6.

4.4 Testing Hypothesis 2: Levels of control in the two treatments

Overall frequencies of control levels are very similar in the two treatments. Averaged over all principals and rounds, frequencies of no control, low and medium control are 15%, 15% and 70% (15%, 18% and 67%) in the Internet (laboratory) treatment. As the session progresses, the medium (low) control level is chosen somewhat more often (less often). In the second half of rounds, frequencies of no control, low and medium control are 13%, 12% and 75% (11%, 18% and 71%) in the Internet (laboratory) treatment. More details on principals' choices are given in Appendix E.7.

We test our second hypothesis with the help of linear mixed models where the dependent variable is the level of control implemented by the principal and random intercepts at the principal and session levels are included. Table 7 reports the estimation results.[23] In model 1 the control level is regressed against an intercept and the experimental condition. Averaged over all principals and rounds, the estimated control level is not significantly different in the two treatments. Moreover, the results of model 2 show that the extent of managerial control increases over time. Model 3 extends model 2 by including differences in beliefs and their interactions with the treatment dummy. Estimation results indicate that the higher the effort increase principals expect when imposing low rather than no control the higher the control level they impose (the coefficient of $b_P(2) - b_P(1)$ is significantly positive). Likewise, the higher the effort increase principals expect when imposing medium rather than low control the higher the control level they impose (the coefficient of $b_P(3) - b_P(2)$ is significantly positive). The beliefs of principals impact their control decisions similarly in the first and in the second half of rounds and the magnitude of the impact is very similar in the two treatments (regression results of model 4). Finally, the regression results

[22] Robustness checks where control aversion is based on effort costs rather than effort levels confirm these findings (see Appendix E.5).

[23] As for the analysis of agents' effort decisions, the estimate of the random intercept at the session level is basically zero for all regression models of the principals' control decisions and the same regressions without a random intercept at the session level generate almost identical standard errors.

are robust to the inclusion of demographic controls and none of the demographics significantly impacts the extent of managerial control (regression results of model 5).

To summarize, the regression results do not support our second hypothesis since the estimated control level is not significantly different in the two treatments. As expected, principals' beliefs significantly influence the level of control they implement. And the impact of beliefs on managerial control is similar in the two treatments.[24]

	Dependent variable: Level of control				
Model	(1)	(2)	(3)	(4)	(5)
Constant	2.528***	2.452***	2.321***	2.332***	2.301***
	(0.049)	(0.052)	(0.047)	(0.048)	(0.077)
Int	0.014	0.010	-0.015	-0.016	-0.022
	(0.067)	(0.071)	(0.065)	(0.066)	(0.067)
$Half2$		0.152***	0.122***	0.092**	0.093**
		(0.036)	(0.035)	(0.041)	(0.041)
$Int * Half2$		0.008	0.003	-0.013	-0.013
		(0.049)	(0.048)	(0.061)	(0.061)
$b_P(2) - b_P(1)$			0.134***	0.123***	0.123***
			(0.020)	(0.025)	(0.025)
$b_P(3) - b_P(2)$			0.135***	0.124***	0.123***
			(0.022)	(0.026)	(0.026)
$Int * [b_P(2) - b_P(1)]$			0.022	0.024	0.024
			(0.027)	(0.032)	(0.032)
$Int * [b_P(3) - b_P(2)]$			-0.010	-0.007	-0.007
			(0.028)	(0.034)	(0.034)
$Half2 * [b_P(2) - b_P(1)]$				0.026	0.025
				(0.036)	(0.036)
$Half2 * [b_P(3) - b_P(2)]$				0.027	0.026
				(0.037)	(0.037)
$Int * Half2 * [b_P(2) - b_P(1)]$				0.010	0.010
				(0.051)	(0.051)
$Int * Half2 * [b_P(3) - b_P(2)]$				0.006	0.006
				(0.051)	(0.051)
Age					0.030
					(0.060)
Male					0.096
					(0.063)
Social					-0.045
					(0.074)
Hum					-0.018
					(0.083)
Tech					-0.055
					(0.089)
Observations	2,200	2,200	2,200	2,200	2,200
Log-likelihood	-2145.224	-2125.083	-2018.572	-2016.659	-2014.685

Notes: Standard errors in parentheses. ***(1%); **(5%); *(10%) significance level.

Table 7: Determinants of the control intensity.

[24] We obtain the same qualitative findings with ordered probit regressions which take the ordinal structure of control levels into account (see Appendix E.8).

Principals' beliefs

Principals correctly expect agents to be intrinsically motivated and they also correctly expect their efforts to increase with control. Averaged over all principals and rounds, expected efforts under no, low and medium control are 2.58, 3.18 and 3.82 (2.93, 3.46 and 4.02) in the Internet (laboratory) treatment. Expected effort, averaged across rounds for each principal, significantly increases with the level of control in each treatment (Wilcoxon matched-pairs signed-ranks tests: $p - value < 0.01$ in all four cases). Since agents' effort differences are significantly different from zero only in the Internet treatment, principals seem unaware that the distance in agency relationships influences how agents' efforts vary with the level of control. In a similar vein, Dickinson and Villeval (2008) conclude that the nature of the agency relationship does not significantly affect the monitoring of their principals.

Figure 4 shows the average effort differences expected by principals over time. Principals expect effort differences to be clearly positive in every round even though actual average effort differences are often close to zero in the laboratory. Monetary payoff-maximizing principals should always enforce medium control (see Figure 2) and on average principals correctly believe that enforcing medium control is the decision which maximizes their monetary payoffs. Principals' monetary payoffs are non-significantly lower in the Internet than in the laboratory treatment. Further details about principals' beliefs and monetary payoffs are given in Appendices E.9 and E.10.

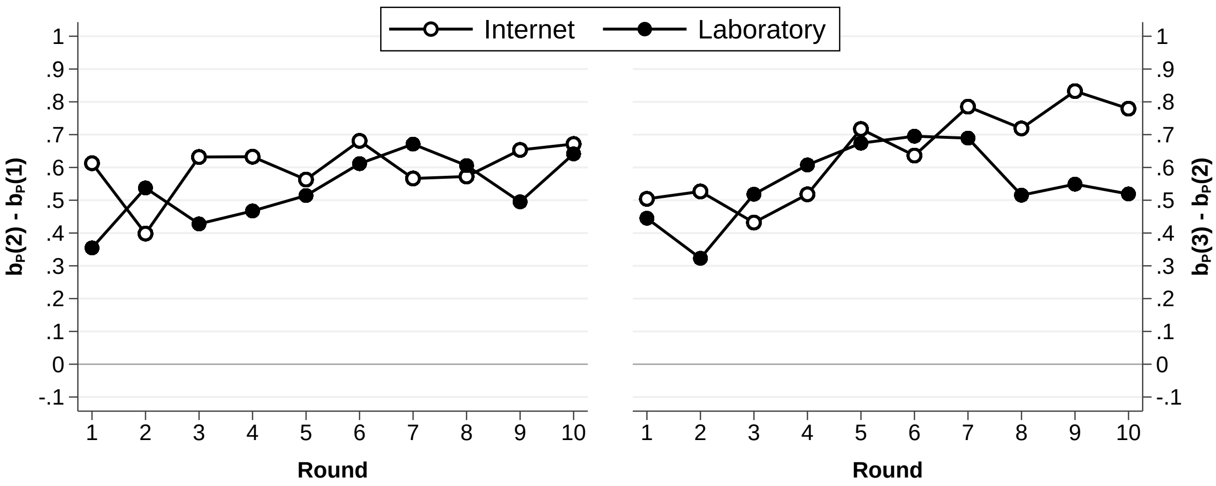


Figure 4: Effort differences expected by principals over time.

5 Conclusion

There is by now conclusive evidence that principals suffer hidden costs of control in a laboratory principal-agent game where the principal can control the agent by implementing a minimum effort requirement before the agent chooses an effort costly to her but beneficial to the principal (Falk and Kosfeld, 2006; Ziegelmeyer, Schmelz, and Ploner, 2012). We report an experiment designed to assess the influence of workplace arrangements on the magnitude of hidden costs of control by comparing the online and laboratory behavior in a straightforward extension of the principal-agent game. Though we rely on the same experimental procedures in both treatments, online subjects enjoy greater anonymity than lab subjects, they interact in a less constrained environment than the laboratory, and there is a larger physically-oriented social distance between them. We find that hidden costs of control are significantly lower online than in the laboratory. The treatment effect is driven by both higher intrinsic motivation and stronger control aversion in the laboratory. Agents' effort differences are fairly stable over time in each treatment which indicates that even experienced agents react more negatively to the implementation of control in the laboratory than over the Internet. We also observe that in both treatments principals choose the highest control level most of the time which maximizes their monetary payoffs.

Our results are of practical relevance for organizations. Though the manipulation of the distance in agency relationships is weak, we find that hidden costs of control are significantly lower in distant than in close agency relationships even for experienced agents. We are therefore reasonably confident that our main qualitative finding generalizes to field agency relationships similar to our experimental ones, implying that work motivation is less likely to be reduced by managerial control in WFH arrangements than in working at the office arrangements. By contrast, the quantitative finding that the benefits of control outweigh the hidden costs of control in each treatment should be taken with a grain of salt. We share the common belief that economic experiments promise qualitative external validity but may not promise quantitative external validity especially when the subject pool consists of students (Kessler and Vesterlund, 2015).

References

Anderhub, V., R. Müller, and C. Schmidt (2001): "Design and evaluation of an economic experiment via the Internet," *Journal of Economic Behavior & Organization*, 46(2), 227–247.

Barmettler, F., E. Fehr, and C. Zehnder (2012): "Big experimenter is watching you! Anonymity and prosocial behavior in the laboratory," *Games and Economic Behavior*, 75, 17–34.

Bartling, B., E. Fehr, and K. M. Schmidt (2012): "Screening, competition, and job design: Economic origins of good jobs," *American Economic Review*, 102(2), 834–864.

Bloom, N., J. Liang, J. Roberts, and Z. J. Ying (2015): "Does working from home work? Evidence from a Chinese experiment," *Quarterly Journal of Economics*, 130(1), 165–218.

Bowles, S., and S. Polanía-Reyes (2012): "Economic incentives and social preferences: substitutes or complements?," *Journal of Economic Literature*, 50(2), 368–425.

Charness, G., E. Haruvy, and D. Sonsino (2007): "Social distance and reciprocity: An Internet experiment," *Journal of Economic Behavior & Organization*, 63(1), 88–103.

Dickinson, D., and M.-C. Villeval (2008): "Does monitoring decrease work effort? The complementarity between agency and crowding-out theories," *Games and Economic Behavior*, 63, 56–76.

Dutcher, G., K. J. Saral, and T. C. Salmon (2015): "Is "real" effort more real?," Working Paper.

Dutcher, G. E. (2012): "The effects of telecommuting on productivity: An experimental examination. The role of dull and creative tasks," *Journal of Economic Behavior & Organization*, 84(1), 355–363.

Eckel, C. C., and R. K. Wilson (2006): "Internet cautions: Experimental games with internet partners," *Experimental Economics*, 9, 53–66.

Eurofound (2012): *Fifth European Working Conditions Survey*. Publications Office of the European Union, Luxembourg.

Falk, A., and M. Kosfeld (2006): "The hidden costs of control," *American Economic Review*, 96(5), 1611–30.

Felstead, A., N. Jewson, and S. Walters (2003): "Managerial control of employees working at home," *British Journal of Industrial Relations*, 41(2), 241–64.

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

Frey, B. S. (1993): "Does monitoring increase work effort? The rivalry with trust and loyality," *Economic Inquiry*, 31(4), 663–670.

Frey, B. S., and R. Jegen (2001): "Motivation crowding theory," *Journal of Economic Surveys*, 15, 589–611.

GfK (2010): "GfK Kaufkraft 2011," http://www.gfk-geomarketing.de/fileadmin/newsletter/pressemitteilung/kaufkraft-deutschland_2011.html, Society for Consumer Research, Accessed: 2015-06-24.

——— (2014): "GfK Purchasing Power Germany 2015," https://www.gfk.com/news-and-events/press-room/press-releases/Pages/purchasing-power-germany-2015.aspx, Society for Consumer Research, Accessed: 2015-06-24.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1(1), 114–125.

Hergueux, J., and N. Jacquemet (2015): "Social preferences in the online laboratory: A randomized experiment," *Experimental Economics*, 18, 251–283.

Kamecke, U. (1997): "Rotations: Matching schemes that efficiently preserve the best reply structure of a one shot game," *International Journal of Game Theory*, 26, 409–417.

Kessler, J. B., and L. Vesterlund (2015): "The external validity of laboratory experiments: The misleading emphasis on quantitative effects," in *Handbook of Experimental Economic Methodology*, ed. by G. R. Fréchette, and A. Schotter, pp. 391–406. Oxford University Press.

Lautsch, B. A., E. E. Kossek, and S. C. Eaton (2009): "Supervisory approaches and paradoxes in managing telecommuting implementation," *Human Relations*, 62(6), 795–827.

Masella, P., S. Meier, and P. Zahn (2014): "Incentives and group identity," *Games and Economic Behavior*, 86, 12–25.

Normann, H.-T., T. Requate, and I. Waichman (2014): "Do short-term laboratory experiments provide valid descriptions of long-term economic interactions? A study of Cournot markets," *Experimental Economics*, 17(3), 371–390.

Riener, G., and S. Wiederhold (2016): "Team building and hidden costs of control," *Journal of Economic Behavior & Organization*, 123, 1–18.

Sliwka, D. (2007): "Trust as a signal of a social norm and the hidden costs of incentive schemes," *American Economic Review*, 97(3), 999–1012.

von Siemens, F. A. (2013): "Intention-based reciprocity and the hidden costs of control," *Journal of Economic Behavior & Organization*, 92, 55–65.

Ziegelmeyer, A., K. Schmelz, and M. Ploner (2012): "Hidden costs of control: Four repetitions and an extension," *Experimental Economics*, 15(2), 323–340.