No. 100

Katrin Schmelz
Anthony Ziegelmeyer

Social Distance and Control Aversion: Evidence from the Internet and the Laboratory

Research Paper Series
Thurgau Institute of Economics and Department of Economics
at the University of Konstanz

Member of
thurgauwissenschaft
www.thurgau-wissenschaft.ch

THURGAU INSTITUTE OF ECONOMICS
at the University of Konstanz

Social Distance and Control Aversion: Evidence from the Internet and the Laboratory*

Katrin Schmelz† and Anthony Ziegelmeyer‡

This version: November 2015

Abstract

We test experimentally whether monitoring is less likely to reduce work motivation in distant than in close principal-agent relationships. Employing the same standard subject pool of students, we compare a laboratory and an internet implementation of an experimental principal-agent game where the principal can impose control at two different levels on the agent. Agency relationships are arguably more distant in the internet than in the laboratory setting. We find that differences in agents' effort due to an increase in the level of control are larger in the internet than in the laboratory experiment. The 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 experiments which indicates that even experienced agents react more negatively to the implementation of control in the laboratory than on the internet.

Keywords: Control; Crowding effects of control; Internet; Motivation; Social distance; Workplace arrangements.

JEL Classification: C81; C90; C93; M52.

1 Introduction

The prevalence of telecommuting in today's organizations is rapidly increasing with a quarter of the European employees mostly working from places other than the office in 2010 (Eurofound, 2012).[1] Like other flexible scheduling and work arrangements, telecommuting challenges existing managerial approaches designed for office employees since such approaches might be inadequate to supervise and elicit performance from remote employees. Lacking any direct observation of employees' effort, employers could either develop supervisory relations based on trust and autonomy or they could turn to more stringent supervisory procedures. Whether different work arrangements call for different managerial approaches heavily

*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. We are grateful to the audiences of invited talks in Bern, Bonn, Cologne, Düsseldorf, Heidelberg, Innsbruck, Jena, Konstanz, Nottingham, Oldenburg and Paris for their stimulating feedback. This project 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]Similar trends are observed in the U.S. (Mateyka, Rapino, and Landivar, 2012) and worldwide (Bloom, Liang, Roberts, and Ying, 2015).

depends on whether the nature of the employment relationship, close or distant, influences employees' reactions towards supervision.

Bruno S. Frey and coauthors have repeatedly argued that the more personal the relationship between employers and employees the more likely monitoring reduces work effort and performance (Frey, 1993, 1997; Frey and Jegen, 2001). The argument distinguishes work situations where the intensity of the personal relationship is strong from work situations where relationships are impersonal. 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 intrinsic motivation. In distant employment relationships intrinsic motivation is less present and crowding out plays little role. Accordingly, the imposition of more stringent monitoring strategies on remote employees are less likely to backfire on employers. This paper reports an experimental test of Frey's hypothesis which sheds light on the situational determinants of control aversion.[2]

We compare 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. The principal-agent game is a straightforward extension of the game implemented by Falk and Kosfeld (2006) in their main treatments. As detailed in Ziegelmeyer, Schmelz, and Ploner (2012), there is by now conclusive evidence that hidden costs of control exist in such a setting. In a laboratory session, the fact that participants have seen each other when entering the laboratory and the physical proximity are likely to create a social bond between participants even if interactions are anonymous. In an internet session, the social bond is likely to be weaker since interactions take place between distant and bodiless participants. Compared to the restrictive laboratory setting, the online setting also reduces the level of external control since participants are free to engage in other activities while making their choices. Both the extent of the social bond and the level of external control differ in office and remote employment relationships, and we use the aggregating term *social distance* to refer to the two aspects of the work arrangement. We postulate that employment relationships are more distant in the internet than in the laboratory experiment, and we explore whether agents' reactions towards control vary with the social distance by comparing agents' efforts in the two experiments.[3]

Our study also addresses several methodological issues related to the measurement of hidden costs of control. First, we move away from the one-shot trial environment implemented by Falk and Kosfeld (2006). We implement a repetitive trial environment in which participants are matched according to the "no-contagion" protocol (Kamecke, 1997).[4] A repetitive trial environment allows participants to gain experience with the interactive situation and experienced behavior in the principal-agent game is likely to be a more reliable measure of control aversion. Since participants are informed about the payoff consequences of their choices after each repetition of the principal-agent game they may learn through personal experience and thought. 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 new aspect of the design enables us to distinguish between the categorical effect of control and the

[2]We will mainly use the term *control aversion* when referring to the crowding-out effect. Control averse agents show a strictly higher effort if not controlled than if controlled. On a sample level, control aversion implies that the principal suffers *hidden costs of control* which increase with the fraction of control averse agents and the relative size of their efforts if controlled and if not controlled.

[3]Our internet experiment is identical to our laboratory experiment except that it is conducted online. In particular, we did our best to vary the social distance without altering the degree of anonymity or the strategic aspects of the interaction (see the details of our experimental design in Section 3).

[4]No participant i is ever matched with a participant who has previously been matched with someone who has been matched with someone participant 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). This matching protocol best preserves the nature of one-shot interactions.

marginal effect of variations in control. Third, to diminish the confounding effect of opportunism on control aversion, we define a payoff function such that expressing the dislike of control bears low costs for selfish agents. Fourth, we elicit and incentivize participants' beliefs about the average behavior of their counterpart which is especially helpful to infer principals' motives for their choices. So, beyond the substantive issue discussed above, our study investigates whether the existence of hidden costs of control is robust to design variations. The results of the two experiments also inform us about the viability of the internet as an alternative to the laboratory for the experimental investigation of control aversion in employment relationships.

Our experimental results can be summarized as follows. First, benefits of control outweigh hidden costs in both experiments. Second, differences in agents' effort due to an increase in the level of control are larger in the internet than in the laboratory experiment and these differences originate from a higher intrinsic motivation and a stronger control aversion in the laboratory. This second result confirms Frey's hypothesis since a smaller social distance induces stronger control aversion. Third, we find that effort differences are fairly stable over rounds in both experiments which indicates that even experienced agents react more negatively to the implementation of control in the laboratory. Fourth, principals choose the highest control level most of the time in both experiments which maximizes their monetary payoff. Yet, principals seem unaware of the Frey (1993) hypothesis as they fail to recognize that the social distance influences how agents' efforts vary with the control level. Finally, we show that the quality of our internet data reaches the high quality of data collected in the laboratory. We conclude that, even though control aversion is weaker on the internet than in the laboratory, the internet is a viable alternative to the laboratory for the experimental investigation of control aversion.

Related literature

Dickinson and Villeval (2008) use a laboratory experiment to analyze the influence of the nature of the agency relationship on the relative importance of the disciplining and crowding-out effects of monitoring on agents' effort. The agent is engaged in a real-effort task which the authors argue is likely to generate substantial intrinsic motivation and the principal chooses the probability with which the agent's output is audited. The agent performs the task after being informed about the principal's monitoring choice. The authors test Frey's (1993) hypothesis by contrasting a distant relationship treatment (stranger matching protocol with preserved anonymity) with a personal relationship treatment (partner matching protocol with face-to-face communication prior to the first interaction). Dickinson and Villeval observe that the disciplining effect dominates the crowding-out effect in both treatments. Still, there is a negative impact of monitoring on performance in the personal relationship treatment but no evidence that the output of agents is crowded-out in the distant relationship treatment. The authors conclude that non-anonymous interpersonal relationships are a major condition for the detrimental effect of monitoring on performance which is consistent with Frey's (1993) hypothesis.[5] Though we agree with Dickinson and Villeval (p. 57) that "it is harder to imagine an interpersonal one-shot relationship", the effect of the nature of the agency relationship is confounded with the effect of reputation in repeated interactions. Our study complements the work of Dickinson and Villeval (2008) in at least two ways. First, our internet experiment implements a highly impersonal agency relationship which provides a stress test for the existence of hidden costs of control. Second, we explore how the influence of the nature of the agency relationship on hidden costs of control evolves when agents gain experience with the interactive situation and get feedback from previous interactions.

[5] Early empirical support for the hypothesis is found in a survey of managers (Barkema, 1995). Being monitored by a CEO had a significant negative effect on hours worked by the manager while being monitored impersonally by a parent company had a significant positive effect.

Numerous experimental studies have investigated whether social preferences are sensitive to the extent of the social bond between participants. In a nutshell, the findings suggest that positive reciprocity increases slightly with social bond (Charness, Haruvy, and Sonsino, 2007), while mixed effects of social bond have been observed with respect to altruism in dictator games (Hoffman, McCabe, and Smith, 1996; Charness and Gneezy, 2008; Hergueux and Jacquemet, 2015) and trust (Glaeser, Laibson, Scheinkman, and Soutter, 2000; Buchan, Johnson, and Croson, 2006; Etang, Fielding, and Knowles, 2011; Hergueux and Jacquemet, 2015). Our study complements this literature by exploring the effect of the social bond on preferences towards control.

In a similar vein, Haley and Fessler (2005) found that subtle cues of the presence of observers increase generosity in dictator games. Eliminating sounds which indicate the presence of others by earmuffs did not affect giving significantly, while displaying stylized eyespots increased generosity. Thus, potential observability or, put differently, potential external control increases prosocial behavior.

A recent small experimental literature on telecommuting has addressed the effect of the environment on individual productivity. Dutcher (2012) compares the effect of the environment (laboratory versus place of choice) on performance in routine and creative tasks and finds that a remote location is detrimental to routine tasks but favorable to creative tasks. Bloom, Liang, Roberts, and Ying (2015) conducted an extensive field experiment with call center employees in a Chinese company. Half of the employees invited to the study volunteered to work from home, and these volunteers were randomly assigned to the treatment group who indeed worked from home and a control group who kept working in the office. Work from home lead to a significant increase in performance of employees who self-selected to work from home. This improvement is driven by more working minutes and higher output per minute. We are not aware of a study that investigates how the location of work interacts with monitoring intensity and our study fills this gap.[6]

Besides our main research question, this study has a substantial methodological aspect as it compares repeated one-shot interactions in the laboratory and on the internet. In their section on related literature, Normann, Requate, and Waichman (2014) review the small literature on repeated-interaction studies comparing behavior in the laboratory and on the internet. They conclude that in experiments with repeated interactions, behavior in these two environments is similar. Note that these earlier studies rely on different subject pools in the laboratory and on the internet. Normann, Requate, and Waichman (2014) employ the same randomized subject pool in both environments to interact repeatedly in a Cournot market and do not find crucial differences either. On the other hand, most studies relying on one-shot interactions do find behavioral differences between the laboratory and the internet. In line with this observation, previous research has shown that on the internet, limited data quality due to an inevitable loss of control compared to the laboratory is of a major concern (Anderhub, Müller, and Schmidt, 2001; Hergueux and Jacquemet, 2015). We invested much effort to minimize this problem. Our careful implementation of repeated trials in the laboratory and on the internet, employing the same standard subject pool of students, allows us to probe the limits of data quality achievable on the internet.

The next section presents the interaction between the principal and the agent that constitutes the basis of our experimental design, and some behavioral predictions are provided. Section 3 outlines the experimental design and procedures of our two experiments and presents detailed hypotheses. Section 4 describes our results. Section 5 discusses the quality of our internet data and Section 6 concludes.

[6] A somewhat related study is Berg, Ghatak, Manjula, Rajasekhar, and Sanchari (2015) who investigate the interaction of social distance and incentive pay in a field experiment with knowledge agents in India. In the absence of incentives, the flow of information is reduced with social distance (measured by the difference in social status). Incentive pay does not affect the dissemination of information in the agent's own social group, but increases the information spread to socially distant groups.

2 Motivation crowding effects of control

We first present the principal-agent interaction that forms the basis of our experimental design. Second, we derive rich behavioral predictions assuming that the agent may be motivated by non-monetary considerations and that control may either substitute for (crowding out) or complement (crowding in) the agent's non-monetary motivations.

2.1 A principal-agent interaction of hierarchical 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 has the possibility to restrict the agent's effort set by choosing one out of three control levels: no control ($\underline{e} = 1$), low control ($\underline{e} = 2$), and 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.

Following Frey's argument, a precondition for control aversion is some degree of intrinsic motivation which can be crowded out. A payoff-maximizing agent might also dislike control but he is unable to express control aversion because his non-monetary motivations are too weak. To diminish the confounding effect of opportunism on control aversion, we define the payoff function such that expressing the dislike of control bears low costs for selfish agents. In the lower range of effort levels, exerting slightly more effort than enforced ($e > \underline{e}$, where $\underline{e} \in \{1, 2, 3\}$) is very cheap for the agent and extremely beneficial for the principal. Moreover, 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. Benefits from the agent's effort are assumed to be concave which reflects productivity losses due to physical restrictions.

2.2 Behavioral predictions

We focus on the agent's behavior and we assume that the principal is interested only in maximizing his own monetary payoff and that she is able to perfectly forecast the agent's effort at any control level. By contrast, the agent's effort is partly driven by non-monetary motivations with the level of control possibly shaping the influence of non-monetary motivations on effort.

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.[7] For a given couple $(\underline{e}, e) \in \{1, 2, 3\} \times \{\underline{e}, \ldots, 10\}$, the agent's utility is

[7] 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.

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

where $\Pi_A(\cdot)$ is the agent's monetary payoff function, $\lambda_0 \geq 0$ measures the agent's non-monetary motivations to exert effort, $\lambda_c \in \mathbb{R}$ measures the categorical effect of control on non-monetary motivations, $\lambda_m \in \mathbb{R}$ measures the marginal effect of variations in control on non-monetary motivations, and the indicator $\mathbb{1}\{\underline{e} > 1\}$ equals 1 if $\underline{e} > 1$ and zero otherwise. In the principal-agent interaction of hierarchical control, non-monetary motivations capture other-regarding preferences that induce agents to help principals 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 non-monetary motivations affect her chosen effort independently of the presence or magnitude of control ($\lambda_c = \lambda_m = 0$). The specification also allows non-monetary motivations 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. Crowding effects may arise because the agent infers from the enforced level of control the motivations of the principal or his beliefs regarding the agent's motivations, or because a restricted set of effort levels undermines the agent's sense of autonomy.[8] For the sake of clarity, 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 she 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 $\lambda_0\left(1+\lambda_c\,\mathbb{1}\{\underline{e}>1\}+\lambda_m\left(\underline{e}-1\right)\right) \leq \underline{e}$;
- $e^*(\underline{e}) \in \{\underline{e}+1, \ldots, 9\}$, a "mid" effort, if $\lambda_0\left(1+\lambda_c\,\mathbb{1}\{\underline{e}>1\}+\lambda_m\left(\underline{e}-1\right)\right) \in \left]\Pi_A(e^*(\underline{e})-1)-\Pi_A(e^*(\underline{e})), \Pi_A(e^*(\underline{e}))-\Pi_A(e^*(\underline{e})+1)\right]$; and
- $e^*(\underline{e}) = 10$, the highest effort, if $\lambda_0\left(1+\lambda_c\,\mathbb{1}\{\underline{e}>1\}+\lambda_m\left(\underline{e}-1\right)\right) > 16$.

Table 2 illustrates the agent's best reply where non-monetary motivations are either weak, moderate or strong, and various crowding effects are considered.

As she knows the agent's best reply, the principal enforces the medium control level in the following three cases: i) Non-monetary motivations are sufficiently weak which leads the agent to exert minimal effort for each level of control enforced by the principal; ii) Non-monetary motivations are 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 non-monetary motivations are 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 non-monetary motivations are sufficiently strong and both $\lambda_c + \lambda_m$ and $\lambda_c + 2\lambda_m$ are sufficiently negative.

[8] Information-based foundations for control-state-preferences in principal-agent interactions of hierarchical control are provided in, among others, Sliwka (2007), Ellingsen and Johannesson (2008), and von Siemens (2013).

	Non-monetary motivations								
	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.

3 Experimental design, procedures, and hypotheses

3.1 Design

In both experiments, participants 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.[9] 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. To diminish the confounding effect of opportunism on agents' reactions towards control, agents' payoffs are such that expressing the dislike of control bears low costs for selfish agents.

In a given session, each participant is assigned either the role of agent or the role of principal. Participants gain experience with the context and the behavior of other participants during 10 rounds of the interaction. Roles are kept constant over all rounds. The matching follows a "no-contagion" protocol which best preserves the nature of one-shot interactions.

Belief elicitation

Before they interact in the employment relationship, participants are asked to guess the average behavior of their counterpart. In each round, participants make three guesses. Principals are asked to guess, for

[9]To avoid demand effects, we did not distinguish between "no control" and a level of control in the instructions and on the decision screens. 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.

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 that control level. Each agent reports his 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 decisions of other participants. Once all guesses and choices have been made in a given round, each participant is only informed about the behavior of his counterpart. Participants do not learn about the correctness of their guess during the experimental session.

Earnings

Each participant 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 experiment. For each participant, depending on the outcomes of random draws, either one of the three guesses is paid or the earnings are determined by the employment 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 participant earns 70 ECUs. Otherwise the participant earns 20 ECUs.

3.2 Practical procedures

Both experiments were conducted with the help of an internet platform developed by the authors and detailed in Appendix A. All 440 participants were students who had agreed to participate in economic experiments. The data were collected in two waves. The first wave was conducted in Jena with 106 students from the University of Jena (November 2010 and January 2011). As we ran only two sessions per experiment in Jena, we conducted a second wave in Konstanz with 334 students from the University of Konstanz participating in another six sessions per experiment (November 2014 and April 2015). All experiments followed the same protocol.[10]

Participants 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 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 participants received a survey token via email. Answering the survey questions took on average 10 minutes and participants had a time frame of a few days to do so. In both experiments, participants completed the survey at their place of choice (e.g. at home).[11]

Participants 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 internet experiment, 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 participants had to log on not later than

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

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

that time. Like for the survey, participants made their interactive choices at their place of choice. The eight sessions of the laboratory experiment took place in the afternoon or evening with a total of 208 participants. Instructions were not read aloud.

In both experiments, participants received their earnings in the laboratory. In the internet experiment, participants 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 experiment, we also added the usual show-up fee to participants' earnings.[12] 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. Participants in Jena (Konstanz) earned 15.42 (19.60) euros on average (about 21 (22) US dollars at the time of the sessions).[13] The participation process is detailed in Appendix A and screens with the instructions are provided in Appendix B.

Participants of the internet and laboratory experiments are very similar with respect to basic socio-demographic characteristics. 52.4% and 49.1% of the participants are female in the laboratory and internet experiment respectively. In the laboratory (internet) experiment, the sample composition according to the field of study is such that 30.8% (33.6%) of the participants belong to the category "business administration & economics", 27.9% (30.6%) belong to the category "other behavioral & social sciences", 18.7% (18.1%) belong to the category "humanities" and 22.6% (17.7%) belong to the category "engineering, life & natural sciences". Participants' age is on average 21.5 and 22.1 years in the laboratory and internet experiments.[14]

3.3 Hypotheses

Our first and main research hypothesis is (a weak version of) the Frey (1993) hypothesis adapted to our experimental environment.

Hypothesis 1 [effort differences]: The difference between the agent's effort under low control ($e(2)$) and the agent's effort under no control ($e(1)$) is strictly larger in the internet than in the laboratory experiment. Similarly, the difference between the agent's effort under medium control ($e(3)$) and the agent's effort under low control ($e(2)$) is strictly larger in the internet than in the laboratory experiment.

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

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

Note that our first hypothesis is consistent with agents expressing (at least) moderate intrinsic motivation in the two experiments and larger control aversion in the laboratory than in the internet experiment, but also with negligible control aversion in the two experiments and moderate intrinsic motivations only

[12] Subjects in the laboratory were paid immediately after the experiment as usual, while internet subjects collected their earnings about three days later, once the official lottery draw had determined their payoffs. While we cannot rule out that time discounting had some influence on the behavior of our internet participants, the nature of this influence is unclear in our interactive setting.

[13] 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 two locations hardly translate into different payoffs in US dollars because of fluctuations in the conversion rates between the two currencies.

[14] In one laboratory session in Jena, two participants dropped out because the internet connection of their clients was interrupted. In the internet experiment, three subjects in Konstanz dropped out. Accordingly, 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 analysis to participants who completed the experiment.

in the laboratory experiment. Since we rely on the strategy method to elicit the agents' chosen efforts, we are able to analyze intrinsic motivations and control aversion separately. Thus, we can decompose hypothesis 1 in its two components, intrinsic motivation and control aversion.

Hypothesis 1a [intrinsic motivation]: In the absence of control, agents' effort is higher in the laboratory than in the internet experiment. Accordingly, we expect higher intrinsic motivation in the laboratory.

$$e(1)\big|_{Internet} < e(1)\big|_{Laboratory}$$

A control averse agent chooses a higher effort in the absence of control than if controlled. More precisely, 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. 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. Similarly, 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.

Hypothesis 1b [control aversion]: The difference between the agent's effort under low control ($e(2)$) and the agent's shifted effort under no control ($max[e(1), 2]$) is negative in the laboratory and strictly greater in the internet experiment. Similarly, the difference between the agent's effort under medium control ($e(3)$) and the agent's shifted effort under low control ($max[e(2), 3]$) is negative in the laboratory and strictly greater in the internet experiment. Put differently, we expect more control aversion in the laboratory.

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

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

In Falk and Kosfeld (2006) principals apparently anticipate the existence of hidden costs of control since a majority of them chooses not to control the agent. In line with this finding and assuming that our first hypothesis is confirmed, we conjecture that agency relationships are characterized by higher effort discretion in the laboratory than in the internet experiment.

Hypothesis 2 [control level]: The control level enforced by the principal is strictly larger in the internet than in the laboratory experiment.

$$\underline{e}\big|_{Internet} > \underline{e}\big|_{Laboratory}$$

Since we elicit the principals' beliefs concerning the agents' chosen efforts, we are able to determine whether principals correctly expect higher effort differences on the internet than in the laboratory.

4 Results

We start by comparing participants' 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 non-monetary motivations and crowding effects at the aggregate level in the two experiments. Finally, we formally test our two research hypotheses.

4.1 Comparing participants' behavior in Jena and Konstanz

Appendix C provides detailed comparisons of the two subject pools with respect to choices and beliefs of participants. The distribution of p-values derived from 42 tests is very close to the uniform distribution (see Appendix C.6). We conclude that behavior in the two locations does not differ in meaningful ways. Any further analysis is based on the pooled data of the two locations.[15]

4.2 Non-monetary motivations and crowding effects

Table 3 shows agents' efforts as a function of the level of control in the two experiments. In each panel, the first row reports the average effort for each control level, the second row reports standard deviation followed by 1st quartile followed by median followed by 3rd quartile for each control level. In both experiments, the average effort in the absence of control is around the largest minimum effort enforceable by the principal which indicates the presence of sizable non-monetary motivations. Additionally, the average effort increases with the control level.[16] Finally, increases in average effort are larger in the internet than in the laboratory experiment.

	No control	Low control	Medium control
Laboratory experiment	**3.46**	**3.51**	**3.63**
(1040 observations)	(2.73;1;2;7)	(1.87;2;2;5)	(1.21;3;3;4)
Internet experiment	**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 assess non-monetary motivations and crowding effects at the aggregate level in the two experiments, we assume that average efforts are best replies of a *representative* agent endowed with control-state-preferences whose monetary payoff function is approximated by $98.5 + 1.5e - 0.75e^2$. At control level $\underline{e}$, the representative agent's best reply equals $\max\{\underline{e}, \min\{1 + \frac{2}{3}\lambda_0 (1 + \lambda_c \mathbb{1}\{\underline{e} > 1\} + \lambda_m (\underline{e} - 1)), 10\}\}$.

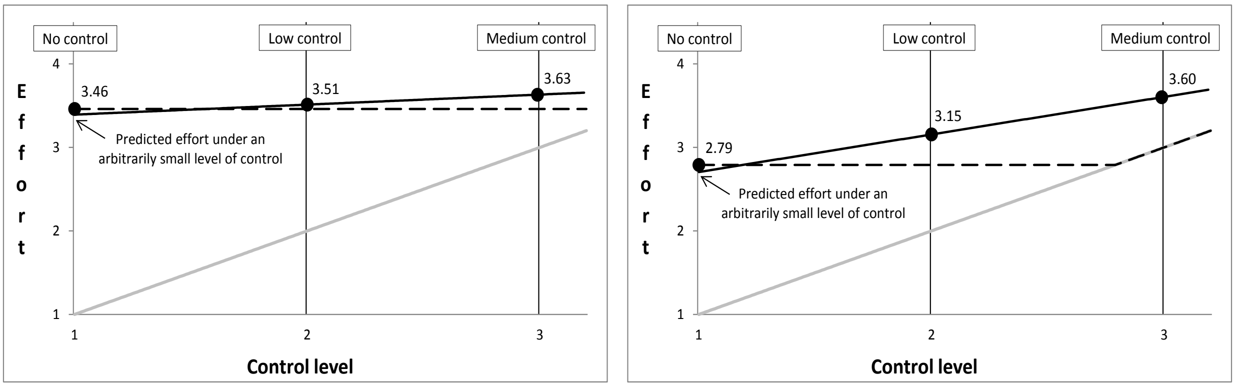
We infer from average efforts in the absence of control that non-monetary motivations are moderate in both experiments and that they are weaker in the internet than in the laboratory experiment ($\hat{\lambda}_0\big|_{Internet} = 2.69$ and $\hat{\lambda}_0\big|_{Laboratory} = 3.69$). Additionally, increases in average effort following an increase in the control level from low to medium control (0.45 and 0.12 in the internet and laboratory experiment respectively) enable us to infer that $\hat{\lambda}_m\big|_{Internet} = 0.25$ and $\hat{\lambda}_m\big|_{Laboratory} = 0.05$. Thus, marginal crowding in is inferred

[15] In a nutshell, we find that: i) in both experiments, 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 experiment, 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 experiment, 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 experiments, 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 experiment, 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 experiment, 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.

[16] If the principal cares solely about his monetary payoffs then she 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.

at the aggregate level in each experiment and the effect is larger in socially distant agency relationships. Finally, given average efforts under low or medium control and the already inferred values of λ_0 and λ_m, the representative agent best replies to an arbitrarily small level of control by exerting effort 2.70 and 3.39 in the internet and laboratory experiment respectively. Control thus categorically crowds out 5 and 3 percent of the effect of non-monetary motivations in the internet and laboratory experiment respectively ($\hat{\lambda}_c|_{Internet} = -0.05$ and $\hat{\lambda}_c|_{Laboratory} = -0.03$).

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 the two experiments. 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 motivated by non-monetary considerations ($\hat{\lambda}_0|_{Internet} = 2.69$ and $\hat{\lambda}_0|_{Laboratory} = 3.69$) 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 (see Appendix D).



(a) Laboratory experiment (b) Internet experiment

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

4.3 Testing hypothesis 1: Effort differences in the two experiments

To formally test whether differences in agents' effort due to an increase in the level of control are larger in the internet than in the laboratory experiment, 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. With this specification we allow the behavior of the same agent in different rounds to be correlated. Table 4 summarizes our estimation results. 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.[17]

In models 1 and 4, the effort difference is regressed against an intercept and the experimental condition where the dummy variable *Int* identifies data from the internet experiment. Averaged over all agents and rounds, the effort difference is always positive but not significantly different from zero in the laboratory experiment and the difference is always strongly significantly increased in the internet experiment. The evidence at the aggregate level reveals that differences in agents' effort due to an increase in the level of control are larger in the internet than in the laboratory experiment which is consistent with Hypothesis 1.

[17]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 behavior of different agents from 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.

The dynamics of effort differences indicate that the behavior of experienced agents is also supportive of our first research 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 of *Int* are significantly positive, while those of *Half2* and *Int*Half2* do not significantly differ from zero. Thus, effort differences and the impact of social distance are fairly stable over time.[18] Estimated coefficients of beliefs 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 summarizes the effort differences over time.

The effects of the experimental condition on effort differences, their dynamics and beliefs differences as detailed in models 2 and 5 are fairly robust with respect to the demographic controls (age, gender and academic major[19]) reported in models 3 and 6. Age and gender do not help to explain effort differences,

[18] 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 in the internet than in the laboratory (χ^2 tests: $p - values < 0.05$).

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

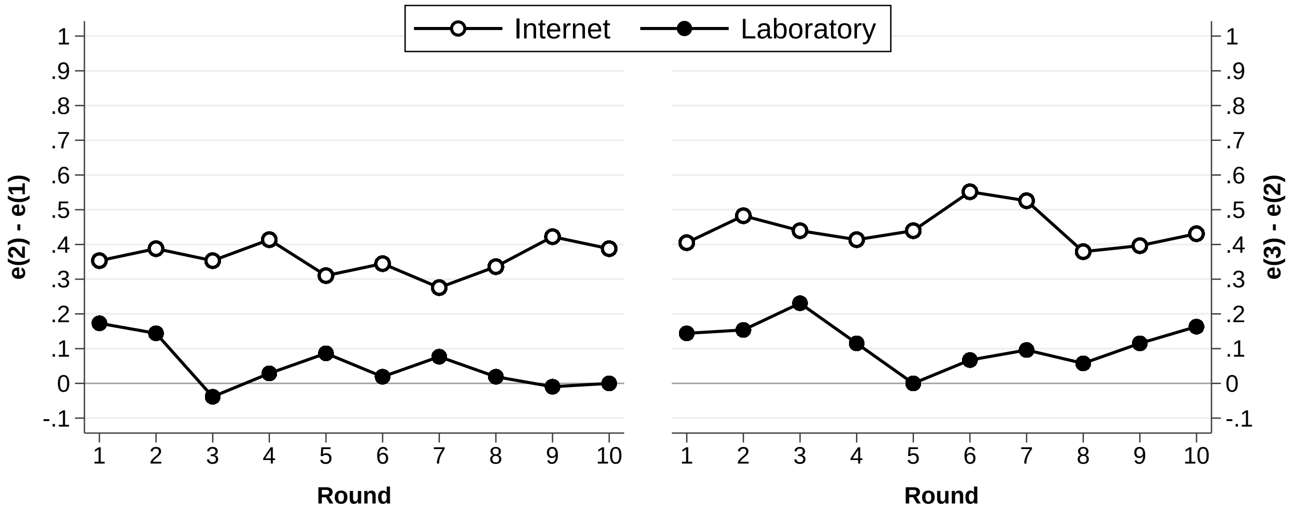


Figure 2: Effort differences over time.

though we find a specific impact of the field of study.[20]

So far, we have tested our first hypothesis by predicting agents' effort differences using linear regression models. Strictly speaking, differences in effort levels are not linear with respect to effort costs for the agent. Thus, we perform robustness regressions on differences in effort costs as the dependent variable. The results are in line with our findings presented in this section as reported in Appendix E.3.

The mere effort differences considered in this section aggregate disciplining and crowding effects of control. As effort differences are largely positive, we infer that disciplining effects outweigh crowding effects in both experiments. Most importantly, the observation that effort differences are larger on the internet than in the laboratory is supportive of the Frey (1993) hypothesis. The next two sections shed light on the mechanisms driving our finding.

Testing hypothesis 1a: Intrinsic motivation in the two experiments

So far, we can conclude that effort differences are supportive of the Frey (1993) hypothesis. To isolate the importance of intrinsic motivation and control aversion, we present a series of additional regression models. Again, we rely on linear mixed models with random intercepts for agents and sessions as reported in Table 5.

In model 1, agents' effort under no control is regressed against an intercept and the experimental condition. Averaged over all rounds, in the laboratory, agents' effort if not controlled exceeds the largest effort level enforceable by the principal and it is significantly lower in the internet experiment. Still, on the internet, agents' effort under no control is highly significantly greater than the minimal effort level of 1 (χ^2 test: $p - value < 0.0000$). We conclude that intrinsic motivation is present in both experiments, but more so in the laboratory.

Model 2 shows that the differences in agents' intrinsic motivation between the two experiments are

[20]The significant impact of $Tech$ on effort differences is not robust when interacting the field of study with the experimental condition. The regression models 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: $p - value = 0.033$), implying that the effort difference $(e(2) - e(1))$ is significantly larger on the internet than in the laboratory in the second half of the experiment.

	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.

fairly stable over time as the coefficients for $Half2$ and $Int * Half2$ do not significantly differ from zero. The impact of beliefs is again either insignificant or small. Figure 3 summarizes agents' average efforts under no control over time.

The effects of the experimental condition on intrinsic motivation, the dynamics and beliefs are fairly robust with respect to our demographic controls reported in model 3. Again, age and gender do not help to explain effort differences, though we find a specific impact of the field of study.[21] Robustness regressions on agents' effort costs under no control confirm our findings as documented in Appendix E.4. Our results reveal that in the absence of control, agents' effort is higher in the laboratory than on the internet, which is in line with higher intrinsic motivation in the laboratory and consistent with Hypothesis 1a.

[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: $p - value = 0.015$), implying that efforts under no control are significantly smaller on the internet than in the laboratory in the second half of rounds.

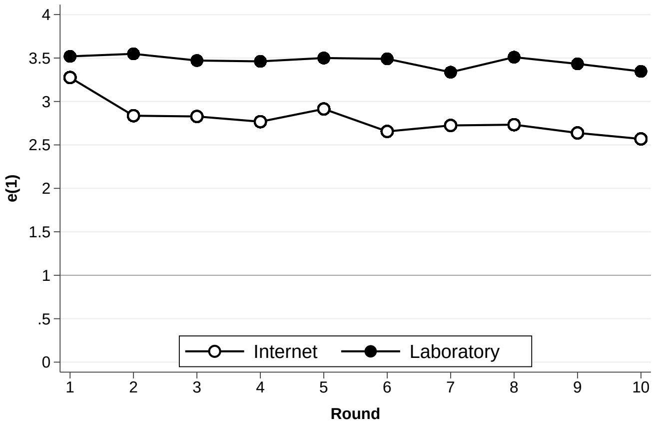


Figure 3: Effort in the absence of control over time.
Remark: Arguably, efforts beyond the minimal requirement of 1 are guided by intrinsic motivation.

Testing hypothesis 1b: Control aversion in the two experiments

Table 6 reports estimation results on control aversion. Models 1 to 6 differ from those in Table 4 only in the dependent variable. Models 1 to 3 predict aversion to low compared to no control, computed by the differences between efforts under low control and the shifted efforts under no control ($e(2) - max[e(1), 2]$), as detailed in Section 3.3. Models 4 to 6 predict aversion to medium compared to low control, reflected by the difference between efforts under medium control and the shifted efforts under low control ($e(3) - max[e(2), 3]$).[22]

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 highly significantly negative in the laboratory experiment and it is always significantly less negative in the internet experiment. Yet, on the internet, aversion to low (medium) control is significantly different from zero at the 5 (10) percent level (χ^2 tests: $p-value = 0.018$ and $p-value = 0.085$). We infer that control crowds out intrinsic motivation in both experiments and but more so in the laboratory than in the internet experiment.

The dynamics of control aversion indicate that the behavior of experienced agents is also supportive of larger crowding out in the laboratory than in the internet experiment. For aversion to both low and medium control, the estimated coefficients of *Half2* are small but (weakly) significantly negative while the coefficient of *Int*Half2* is insignificant. Thus, experienced agents tend to express somewhat more control aversion than naive agents in both experiments. Estimated coefficients of beliefs differences hardly differ from zero. Figure 4 summarizes control aversion over time.

The impact of the experimental condition on control aversion, their dynamics and beliefs differences as detailed in models 2 and 5 are fairly robust with respect to the demographic controls reported in models 3 and 6. Age and gender do not help to explain control aversion, and we do not find a consistent impact of the field of study on control aversion. Robustness regressions in which control aversion is computed based on effort costs instead of efforts support the main results reported in this section (see Appendix E.5). Accordingly, our finding that control aversion is more negative in the laboratory than on the internet is consistent with Hypothesis 1b.

In a nutshell, the data confirm Hypothesis 1 since agents' effort differences are strictly larger on the internet than in the laboratory. This is driven by both higher intrinsic motivation (Hypothesis 1a) and stronger control aversion (Hypothesis 1b) in the laboratory than on the internet. Note that the effect of

[22]In fact, our measure of control aversion is conservative as it comprises crowding out as well as crowding in.

	Dependent variable: Aversion to					
	low (vs. no) control $e(2) - max[e(1), 2]$			medium (vs. low) 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.

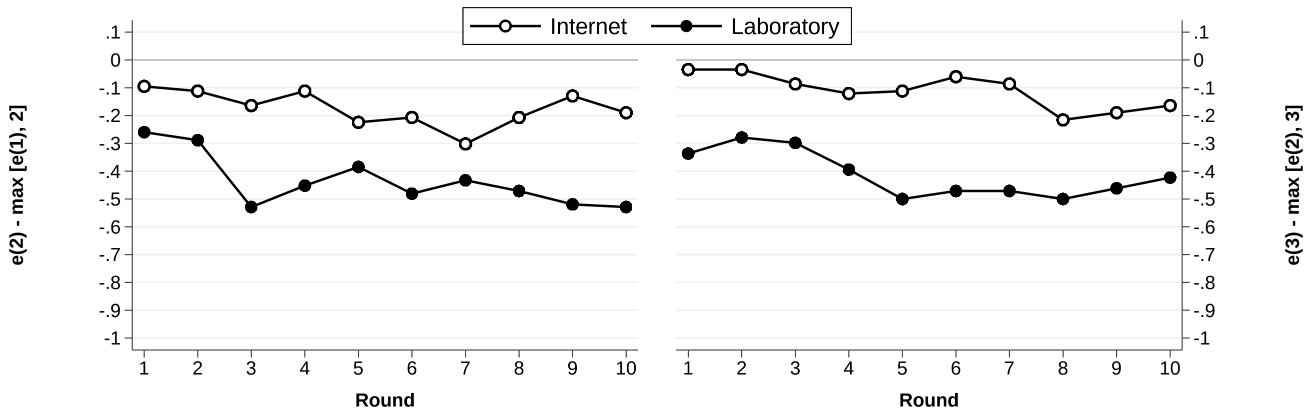


Figure 4: Control aversion over time.

the social distance on effort differences and control aversion is similar at the low and at the medium control level. This observation suggests that the social distance has a strong influence on marginal crowding out and that categorical crowding effects are small in both experiments.[23]

Agents' beliefs

Agents correctly expect little effort discretion in both experiments. 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) experiment. We do not reject the null hypothesis that the distribution of individual beliefs, averaged across rounds for each agent, concerning the frequency of principals who choose the no, low or medium control level is the same in the two experiments at the 10 percent level (Wilcoxon rank-sum test: $p-values > 0.1$). Differences between agents' beliefs in the two experiments 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 both experiments. 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) experiment.

To further 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 rounds, we do not reject the null hypothesis that the distribution of the mean squared difference is the same in the two experiments (Wilcoxon rank-sum test: $p-value > 0.1$). In the second half of rounds, the accuracy of agents' beliefs does not significantly differ between the two experiments either (Wilcoxon rank-sum test: $p-value > 0.1$). Experienced agents do not predict principals' control decisions significantly better in any of the experiments, and they expect agency relationships to be characterized by slightly higher effort control in the laboratory than on the internet, though not significantly so. Further details about agents' beliefs are given in Appendix E.6.

4.4 Testing hypothesis 2: Levels of control in the two experiments

Overall frequencies of control levels in the two experiments are very similar. 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) experiment. As the session progresses, the medium (respectively low) control level is chosen somewhat more often (respectively 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) experiment. More details on principals' choices are given in Appendix E.7.

To test our second hypothesis, we conduct a formal econometric analysis of the principals' control decisions. Again, we consider a series of regression models estimated by linear mixed effects models where random intercepts at the principal and session levels are included. Table 7 summarizes our estimation results.[24]

In model 1 the control level is regressed against an intercept and the experimental condition where the dummy variable *Int* identifies data from the internet experiment. Averaged over all principals and rounds, the control level is not significantly different between the internet and the laboratory experiment. In both environments, agency relationships are characterized by significantly higher effort control in the second half of rounds (model 2).

As expected, principals' beliefs influence their decisions. Model 3 extends model 2 by including differences in beliefs and their interactions with the experiment. Estimation results indicate that the

[23]The effort difference $(e(3)-e(2))$ reflects the marginal effect of control, whereas the effort difference $(e(2)-e(1))$ includes both a categorical and a marginal effect of control.

[24]Similarly to 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.

	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.

higher the effort increase principals expect due to the imposition of low compared to 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 due to the imposition of medium compared to low control, the higher the control level they impose (the coefficient of $b_P(3) - b_P(2)$ is significantly positive). Moreover, principals' beliefs impact their decisions similarly in the first and in the second half of rounds (model 4) and the magnitude of the impact is very similar in the two experiments.

The results are robust with respect to our demographic controls which do not help predicting the principal's choice (model 5). Our findings are also robust in ordered probit regressions which take the ordinal structure of control levels into account (see Appendix E.8). Clearly, the evidence does not support our second hypothesis.

Principals' beliefs

Principals correctly expect agents to be motivated by non-monetary considerations and they also correctly expect effort 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) experiment. Expected effort, averaged across rounds for each principal, significantly increases with the level of control in each of the two experiments (Wilcoxon matched-pairs signed-ranks tests: $p - value < 0.01$ in all four cases). Given the fact that agents' effort differences are significantly different from zero only on the internet but not in the laboratory (see Table 4 in Section 4.3), principals do not seem to take social distance into account.

Figure 5 shows the average effort differences expected by principals over time. Principals expect effort differences to be strictly positive in every round even though actual average effort differences are occasionally close to zero particularly in the laboratory experiment. Monetary payoff-maximizing principals should always enforce medium control (see Figure 2), and on average principals correctly believe that enforcing medium control is always the decision which maximizes their monetary payoffs. Principals' monetary payoffs are lower in the internet than in the laboratory experiment but, according to regression results, not significantly so. In line with decreasing agents' average efforts (see Figure 10 in Appendix E.1), principals earn less over time in both experiments and this difference is significant only on the internet. Further details about principals' beliefs and monetary payoffs are given in Appendices E.9 and E.10.

In sum, principals are not aware that the social distance influences how agents' efforts vary with the level of control. Dickinson and Villeval (2008) derive a similar conclusion by noting that the nature of the agency relationship does not significantly affect the principals' behavior.

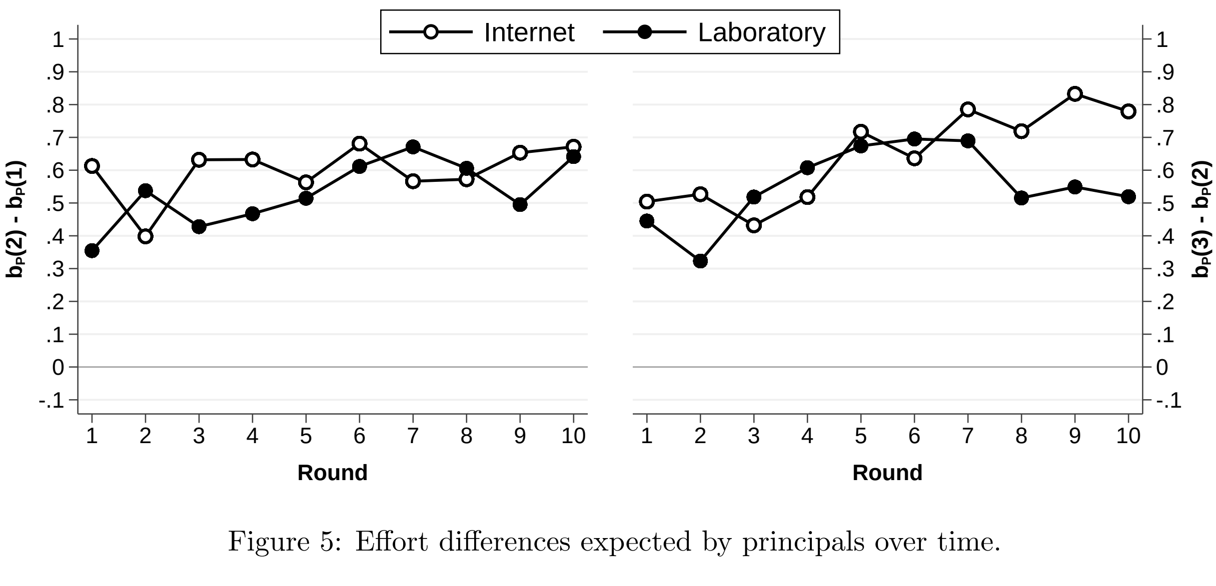


Figure 5: Effort differences expected by principals over time.

5 Data quality on the internet

Due to the inevitable loss of control, internet data might be of lower quality than laboratory data (Anderhub, Müller, and Schmidt, 2001; Hergueux and Jacquemet, 2015). We implemented a number of measures to minimize the difference in the data quality of our two experiments as detailed in Appendices A and B. In particular, decision times have been reported to be lower on the internet than in the laboratory (Anderhub, Müller, and Schmidt, 2001; Shavit, Sonsino, and Benzion, 2001; Hergueux and Jacquemet, 2015) and they seem related to the pro-sociality of choices (for a recent discussion, see Krajbich, Bartling, Hare, and Fehr, 2015). This might be of concern for our study since the non-monetary motivations of our agents determine their level of intrinsic motivation in the employment relationship. In an attempt to eliminate the confound of potentially different decision times, we implemented a read mode which prevents participants from just clicking through the screens. In the read mode, all active items like buttons or input elements were locked. Only after a few seconds, the screen switched into an edit mode where items became unlocked for participants to enter their decisions and continue.[25]

We assess the extent of data quality loss on the internet through several comparisons. First, we analyze whether subjects on the internet had more difficulties to answer the control questions than subjects in the laboratory. We recorded the number of trials a subject required to answer the control questions. In the laboratory (internet) experiment, 64% (61%) of subjects succeeded to answer all questions correctly after the first attempt, only 7% (6%) had to try more than twice, and 1% (1%) failed to find the correct solutions after three attempts.[26] We fail to reject the null hypothesis that the distribution of the number of trials is identical in the two experiments (Wilcoxon rank-sum test: $p - value = 0.524$). Second, we compare the frequency of confused choices in the two experiments. We argue that an effort level chosen by the agent which gives her a lower payoff than the principal indicates confusion (effort levels greater than the fair effort level of 7). Such choices are rare in both experiments. Considering all three choices in all ten rounds, 14 (13) agents make a confused choice at least once in the laboratory (internet) experiment, while 87% (89%) never choose an effort beyond 7. Our data do not suggest that confused choices would occur more often on the internet, and they are almost extinguished in the final round in both experiments as evident from Figure 11 in Appendix E.1. Third, we find that the variance of agents' choices under no and low control tends to be higher in the laboratory than on the internet (see Table 3). Thus, contrary to Anderhub, Müller, and Schmidt (2001) or Shavit, Sonsino, and Benzion (2001) but in line with Hergueux and Jacquemet (2015), we do not find that behavior is more dispersed on the internet. Notice that the higher variance of agents' choices in the laboratory is likely to be explained by a higher intrinsic motivation.

Most online studies are implemented in a one-shot trial environment. Drawing on previous discussions on a loss of control and more confusion on the internet, we argue that repetitions are even more important in online experiments than in the laboratory, and we implemented a repetitive trial environment which allows participants to gain experience with the interactive situation. As opposed to Normann, Requate, and Waichman (2014, and the references therein), we find behavioral differences between the laboratory and the internet in a repetitive trial environment. Concretely, differences in agents' effort due to an increase in the level of control are larger in the internet than in the laboratory experiment. These differences are stable over time and cannot be attributed to more confusion on the internet. We conjecture

[25]On the decision and belief screens of round 1, the read mode lasted for 15 seconds and screens were unlocked after 5 seconds in later rounds.

[26]After three trials, subjects continued with the experiment even if they had not answered all questions correctly. We favored this way instead of excluding them from further participation because we were reluctant to let our subject pool be harmed by frustration. After the third trial, two (three) subjects in the laboratory (internet) experiment had not answered all questions correctly. Four of them answered at least four out of the six questions correctly, and none of them got all questions wrong. The correct solutions were shown to subjects who had entered false answers (see the screenshots in Appendix B.3).

that in studies with more noise due to more complex designs or non-student samples who have to get used to the abstract experimental setting, repetitions are highly valuable as confusion is likely to be reduced over time.

We conclude that, thanks to a careful implementation of the employment relationship and the use of the same subject pool in the two experiments, data of similar quality were collected on the internet and in the laboratory.

6 Conclusion

Our study sheds light on the situational determinants of control aversion by challenging Frey's (1993) hypothesis that the more personal the relationship between employers and employees, the more likely monitoring reduces work effort and performance. We compare an internet and a laboratory implementation of an experimental principal-agent game where the principal can impose control at two different levels on the agent. Agency relationships are highly impersonal in both environments, and arguably less personal in the internet than in the laboratory setting. Thus, our experimental setup enables us to perform a stress test on Frey's hypothesis.

We find that benefits of control outweigh hidden costs of control in both experiments, which is not surprising since interactions are highly impersonal in our two settings. Consistent with Frey's hypothesis, we also show that differences in agents' effort due to an increase in the level of control are larger on average in the laboratory than in the internet experiment. The effect is driven by both stronger non-monetary motivations and stronger control aversion in the laboratory than in the internet experiment. Principals choose the highest control level most of the time in both experiments which maximizes their monetary payoff. Yet, principals seem unaware of Frey's hypothesis as they fail to recognize that the social distance influences how agents' efforts vary with the control level.

Our findings are in agreement with those of Dickinson and Villeval (2008) which suggests that the effect of the social distance on control aversion is robust. Like them, we confirm Frey's hypothesis, we find that principals are unaware of it, and we observe that the disciplining effect dominates the crowding out effect in both experiments.

Our findings also complement the insights of Dickinson and Villeval in important ways. First, they conclude that non-anonymous interpersonal relationships are a major condition for the detrimental effect of monitoring on performance. Contrary to them, we find a negative impact of control on performance in anonymous laboratory relationships in our setting, and control aversion even occurs in highly impersonal interactions on the internet. Accordingly, even in highly impersonal agency relationships the imposition of control on agents might be unprofitable to principals.

Second, an interesting feature of our design is that the principal can control the agent by imposing either a low or a medium effort level before the agent chooses an effort. We show that the effect of the social distance on control aversion is similar at the low and at the medium control level. This result suggests that the social distance has a strong influence on marginal crowding out and categorical crowding effects are small in both experiments.

Finally, another interesting feature of our design is the possibility to explore how the effect of the social distance on control aversion varies when agents gain experience with the interactive situation. We show that effort differences are fairly stable over rounds in both experiments which indicates that average behavior is robust over time and not only naive but also experienced agents react more negatively to the implementation of control in the laboratory. Moreover, experienced agents tend to express more control aversion than naive agents in both experiments.

The experimental test of Frey's hypothesis aside, our study has interesting methodological aspects as it compares crowding out in repeated one-shot interactions in the laboratory and on the internet. Though control aversion is more prevalent in the laboratory than in the internet experiment, the fact that it occurs in both environments leads us to conclude that the internet is a viable (though conservative) alternative to the laboratory for the experimental investigation of control aversion. Moreover, the behavioral differences between the two settings are stable over time and cannot be attributed to more confusion on the internet. Thus, we were able to obtain data of very similar quality on the internet as in the laboratory. Nonetheless, we conjecture that in studies with more noise due to more complex designs or non-student samples who have to get used to the abstract experimental setting, repetitions are highly valuable as confusion is likely to be reduced over time.

Our results are of practical relevance for organizations. The implementation of social distance in our design, affecting both the social bond between employer and employee as well as the level of external control, captures crucial features of telework as opposed to office work. We find that control has weaker negative effects on motivation on the internet than in the laboratory, implying that remote employees are more willing to accept being monitored than office employees. Since our manipulation is weak, we expect the effects to be even stronger in the field.

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.

Barkema, H. G. (1995): "Do top managers work harder when they are monitored?," *Kyklos*, 48(1), 19–42.

Berg, E., M. Ghatak, R. Manjula, D. Rajasekhar, and R. Sanchari (2015): "Motivating knowledge agents: Can incentive pay overcome social distance?," Working Paper. Available at `http://www.erlendberg.info/agentincentives.pdf`.

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.

Buchan, N. R., E. J. Johnson, and R. T. A. Croson (2006): "Let's get personal: An international examination of the influence of communication, culture and social distance on other regarding preferences," *Journal of Economic Behavior & Organization*, 60(3), 373–398.

Charness, G., and U. Gneezy (2008): "What's in a name? Anonymity and social distance in dictator and ultimatum games," *Journal of Economic Behavior & Organization*, 68(1), 29–35.

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. 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.

ELLINGSEN, T., AND M. JOHANNESSON (2008): "Pride and prejudice: The human side of incentive theory," *American Economic Review*, 98(3), 990–1008.

ETANG, A., D. FIELDING, AND S. KNOWLES (2011): "Does trust extend beyond the village? Experimental trust and social distance in Cameroon," *Experimental Economics*, 14(1), 15–35.

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.

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.

GLAESER, E. L., D. I. LAIBSON, J. A. SCHEINKMAN, AND C. L. SOUTTER (2000): "Measuring trust," *Quarterly Journal of Economics*, 115(3), 811–846.

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

HALEY, K. J., AND D. M. T. FESSLER (2005): "Nobody's watching? Subtle cues affect generosity in an anonymous economic game," *Evolution and Human Behavior*, 26(3), 245–256.

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

HOFFMAN, E., K. MCCABE, AND V. L. SMITH (1996): "Social distance and other-regarding behavior in dictator games," *American Economic Review*, 86(3), 653–660.

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.

KRAJBICH, I., B. BARTLING, T. HARE, AND E. FEHR (2015): "Rethinking fast and slow based on a critique of reaction-time reverse inference," *Nature Communications*, 6, 7455.

MATEYKA, P. J., M. A. RAPINO, AND L. C. LANDIVAR (2012): *Home-based workers in the United States: 2010*. U.S. Census Bureau, Current Population Reports.

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.

Shavit, T., D. Sonsino, and U. Benzion (2001): "A comparative study of lotteries-evaluation in class and on the Web," *Journal of Economic Psychology*, 22(4), 483–491.

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.

THURGAU INSTITUTE
OF ECONOMICS
at the University of Konstanz

Hauptstr. 90
CH-8280 Kreuzlingen 2

Telefon: +41 (0)71 677 05 10
Telefax: +41 (0)71 677 05 11

info@twi-kreuzlingen.ch
www.twi-kreuzlingen.ch