Management Practices and Employee Performance -
Causal Evidence from Field- and Laboratory Experiments

Inauguraldissertation
zur
Erlangung des Doktorgrades
der

Wirtschafts-und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2020

vorgelegt
von
Timo Vogelsang, M.Sc.
aus
Wuppertal

Referent: Professor Dr. Dirk Sliwka
Korreferent: Professor Dr. Matthias Heinz
Tag der Promotion: 07.05.2020

Für meine aktuelle und meine zukünftige Familie

Contents

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 1

1. Introduction

“Incentives are the essence of economics” (Prendergast 1999, p.7). With this claim starts one of the most cited paper in the literature on incentives in organizations. The importance of incentives in organizations arises due to unaligned interests of employees (agents) and employers (principals). The resulting moral hazard problem and possible solutions in terms of explicit incentives have been in the focus of the theoretical literature for decades (see, e.g., Jensen and Meckling 1972, Holmström and Milgrom 1991, Baker 1992). Most of the empirical literature has long relied on observational data without much exogenous variation and thus only provided correlational evidence (for an overview see, e.g., Prendergast 1999). With their seminal papers on performance pay, Lazear (2000) and Banker et al. (2000) started the causal empirical literature on the role of employee incentives to increase performance and to act in the interest of the employer. Especially the notion of causality when investigating performance effects has become increasingly important ever since.

Many studies in the last 20 years make use of field experiments (randomized controlled trials), exploiting exogenous variation to establish causal relationships and reduce correlational evidence. Experiments on performance incentives are summarized in Bandiera et al. (2011), List and Rasul (2011), Levitt and Neckermann (2015), and Lazear (2018). The general notion of large parts of this literature can be summarized with performance incentives having a positive effect on performance. Most of the early literature, however, investigates the effects in rather rare and specific settings that might not be directly applied without further investigation to larger organizations in developed countries. They use simple and repetitive tasks for which an accessible measure of effort exist, temporary workers, or settings in which effort is not easily to monitor. These settings possibly provide a good environment for performance incentive to work (as argued, for instance, by Prendergast 1999). Only few studies (see, e.g., Delfgaauw et al. 2013, Friebel et al. 2017) test performance incentives in larger organization in developed countries with more complex working tasks and arguably different environments.

Importantly, performance incentives are not the only management practice to align interests between employees and employers. Scholars in the accounting literature specifically distinguish between the decision-facilitating and decision-influencing role of information (Demski and Feltham 1976, Baiman 1982). Both types of information can be used to align interests in organizations. Providing employees with decision-facilitating information to reduce their uncertainty about specific tasks can direct the employees towards favorable activities for the employer. Using decision-influencing information to set an incentive or to use the information for monitoring also aligns the interests as explained above.

Moreover, organizations continuously try to develop novel tools to attract the interest of employees and generate additional performance increases. Here, leisure time has been increasingly discussed in the last years because of an increased preference for more time-off (see, e.g., Twenge et al. 2010). Companies not only allow for flexible working hours and home office but also try to substitute wage increases with increases in vacation days (see examples of the German railway organization Deutsche Bahn in 2016 or the German labor union IG Metall in 2018). Despite its interesting features, the time domain is only very rarely studied in the accounting, economics and management literature as a medium for rewards or exchange.

This dissertation combines and addresses the three paragraphs above and advances our understanding in these fields. Moreover, it contributes to the field methodologically. First, two of the three chapters use field experiments within a large German retailer to gain causal insights on specific management practices and management controls. While the use of field experiments is nowadays not novel anymore, implementation of field experiments inside big modern organizations is still scarce. Second, both field experiments make use of additional data gathered throughout the experiments with surveys and interviews to investigate the underlying channel of the effect and changes in agents' behavior. While ex-post rationalizations might have their limitations, I believe that our studies and the understanding of the estimated effects benefited a lot from this procedure. Third, in the last chapter I make use of a laboratory experiment to obtain data that allows me to precisely estimate the treatment effect and underlying mechanism. To obtain further external validity for the results, I run a large online survey among German HR-managers. This combination of laboratory data and survey data from the field should reduce skepticisms concerning the artificial environment of laboratory experiments.

Chapter Two is based on Manthei, Sliwka, and Vogelsang (2018).[1] In the second chapter, we study the traditional management practice of monetary performance pay in a large discount retail chain in Germany.

[1] Chapter two is joint work with Kathrin Manthei and Dirk Sliwka. The experiment was preregistered at the AEA RCT registry (AEARCTR-0000961 and AEARCTR-0001758). The idea of this study, the design of the intervention as well as the final paper was joint work. The implementation, data analysis, and writing of the first draft was mainly my work. Dirk Sliwka developed the theoretical model.

In this study, we report the results of two field experiments in a retail chain and find that the effectiveness of performance pay crucially hinges on prior job experience. We first introduce sales-based performance pay for district managers and estimate negligible treatment effects. After conducting phone interviews, we became aware of the fact that store managers might be more suitable for the sales-based performance pay as they might have more influences on sales. Therefore, we run a second experiment exactly one year later incentivizing the store managers of the same region with the exact same incentive on the same KPI. In a third group, we further simplified our normalization. Again, we find negligible treatment effects. Based on more surveys and interviews we develop the idea that the environment can possibly be characterized by learning-by-doing. Many store managers stated in the survey that the simply cannot do much more to increase the KPI further because they already did a lot in the past.

We then formalized this idea demonstrating that the effect of performance pay decreases with experience and may even vanish in the limit. We provide empirical evidence in line with this hypothesis. We create a proxy for experience from the tenure of store managers, age of store managers and age of the store. Redoing our analysis, we find (only) positive treatment effects in stores with low job experience.

This study adds evidence from a firm-level field experiment on monetary incentives to influence managers' decision and increase their performance to the existing literature. In contrast to most of the previous studies, we do not find an average performance effect. It further contributes to literatures on learning-by-doing and habit formation. Moreover, it has practical implications for organization who are about to implement monetary incentives for their employees. It seems to be important that the incentivized variable still has some possibility to increase. Moreover, one might think about incentivizing managers especially early in their careers or changing the incentivized key figure from time to time.

Chapter Three is based on Manthei, Sliwka, and Vogelsang (2019).[2] After the two performance pay experiments described above, we came to the conclusion that we would need a different key figure in case we want to set a useful performance incentive and also to study further management practices. We decided to introduce a simplified profit metric for the stores (*profit = gross margin – personnel costs – inventory losses*). This profit metric covers all

[2] Chapter three is joint work with Kathrin Manthei and Dirk Sliwka. The experiment was preregistered at the AEA RCT registry (AEARCTR-0002127). The idea of this study, the design of the intervention as well as the final paper was joint work. The implementation, data analysis, and writing of the first draft was mainly my work. Dirk Sliwka developed the theoretical model.

components that can be influenced by the store manager and was never in the focus as an aggregated measure before.

In this third chapter, we eventually study the core role of managerial accounting - the processing and provision of information to facilitate managers' decisions and the influence of their behavior through incentives. We study the impact of these two roles of information on profits implementing a field experiment in the same large retail chain. In a 2x2 factorial design we vary: (i) whether store managers obtain access to decision-facilitating accounting information to raise profits and (ii) whether they receive performance pay based on an objective profit metric to influence their decisions. We facilitate the managers profit decisions by disclosing individual product margin on an aggregated level to them. Specifically, we categorize products in five categories depending on their margins and display the categories on managers handheld devices whenever they scan a product. We further designed an online training video to brush up their knowledge on how to influence store profits and to explain the newly developed product margins to them. The other intervention implements performance pay for increases in the simplified profit metric to study the effect of decision-influencing.

Based on a formal model we hypothesize that both practices, decision-facilitating and decision-influencing, are complements. We find that both practices indeed increase profits. In contrast to our hypothesis, we find no evidence of a complementarity. In fact, approximately 75% of the combined effect can be explained by the decision-facilitating intervention. Net of bonus costs, the mere provision of decision-facilitating information even tends to outperform the combined intervention.

Data from online questionnaires reveal that we indeed changed the managers' behavior by providing decision-facilitating information. Once provided with the additional information, managers in these groups focus more on product placements and ordering compared to the groups without the margin information (control group and solely bonus group). Detailed financial data on sales volume and amount of products sold depending on margin categories further support this finding. Here we find that managers with the margin information indeed focus more on high margin products.

The results contribute to the broad managerial accounting literature on influencing and facilitating managers' decisions. Previous studies have focused mainly on the single effects of both roles of managerial accounting information. Our study gives the opportunity to estimate causal performance effects of both roles as well as their interplay within the same firm. We can precisely compare the effects and, furthermore, investigate behavioral changes besides the pure effect on financial KPI's. It also has practical implications for organizations as it seems that if employees have some aligned interest with the employer (some incentive to use the decision-facilitating information) the provision of simple information to help employees in their decision-making can have performance effects of similar magnitude as monetary incentives but come with much lower costs.

Chapter Four is based on Vogelsang (2019).[3] The last chapter shows effects of a novel performance incentive to motivate employees. The previous chapters focus on the standard approach of a monetary bonus to increase employees' performance. In fact, previous studies in all kind of research areas mainly focused on money as a medium of reward and exchange. This is interesting as time is a resource which is of similar importance for humans as money is and further has some interesting characteristics. Moreover, it seems that nowadays time becomes increasingly important in current labor markets (see, e.g., Mas and Pallais 2017, Wiswall and Zafar 2018, Katz and Krueger 2019).

In this fourth chapter, I study how a gift of more leisure time affects employees' performance in a real-effort laboratory experiment. A possible theoretical idea could be that additional leisure time of-the-job reduces the marginal utility of leisure on-the-job and thus employees are more productive because they consume less leisure while working. Employees in this laboratory experiment have to work on a very easy computerized task with the opportunity to browse the internet at any time they want. Results show that a monetary gift of a 75% wage increase does not alter employees' performance, compared to a baseline of no gift. A comparable gift of more leisure time, however, significantly increases employee performance by 25% during the employees' shared working time. The data obtained from this laboratory experiment gives me further the opportunity to precisely study the mechanism for this effect. In principle, employees in the laboratory could work faster or browse the Internet less often. The mechanism for the performance difference is a significant reduction in on-the-job leisure (Internet) consumption by 45% if employees can leave the laboratory earlier. An online survey experiment among human resource managers provides external validity. Managers anticipate the mechanism of on-the-job leisure reduction and point to possible further advantages of leisure time gifts over monetary gifts.

The results of this chapter contribute to the literature of unconditional bonuses (gifts), distortionary behavior during working hours and possibly also to the research on shorter working hours. It shows how the usage of different compensation domains can lead to surprising effects and that a reduction in working time does not necessarily mean a loss in

[3] Chapter four is my single-authored paper.

production. Practically, it adds an interesting reward domain to the possible set of compensation and incentive opportunities for organizations with notable characteristics that might also be highly appreciated in the current labor market.

The research presented in this thesis contributes to the incentive literature in accounting, economics and management. It evaluates performance pay as an existing and popular managements practice/ management control from different perspectives and shows that other ways exist that can lead to performance increases and might be more cost efficient. Moreover, it contributes to the literature the first evaluation of leisure time as a novel bonus domain. Lastly, the research in this thesis contributes methodologically to the empirical literature in accounting, economics and management. It provides insights on how design a field experiment (RCT) within a large firm in a developed country. The studies moreover show the importance of gathering additional data when conducting field experiments to get to the precise mechanisms and behavioral changes. It also gives credit to laboratory experiments as a complement to field experiments. For research questions for which precise and maybe even sensible data are important, as in the research presented above, laboratory experiment, although arguably artificial, still have their importance. I further deeply believe that the combination of laboratory data together with survey data from the field will help the field to earn more prestige in the future.

To conclude, this dissertation consists of three separate articles that answer questions in the area of management practices, motivation and incentives. Likewise, the dissertation raises new questions and provides evidence for the fact that we still do not know a lot for sure (although we might think we do).

1.1 References of Chapter 1

Baiman, Stanley. “Agency research in managerial accounting: a survey.” *Journal of Accounting Literature* 1 (1982): 154-213.

Baker, George P. “Incentive Contracts and Performance Measurement.” *Journal of Political Economy* 100.3 (1992): 598-614.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Field experiments with firms." *Journal of Economic Perspectives* 25.3 (2011): 63-82.

Banker, Rajiv D., Seok-Young Lee, Gordon Potter, and Dhinu Srinivasan. “An Empirical Analysis of Continuing Improvements Following the Implementation of a Performance-Based Compensation Plan.” *Journal of Accounting and Economics* 30.3 (2000): 315-350.

Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke. "Tournament incentives in the field: Gender differences in the workplace." *Journal of Labor Economics* 31.2 (2013): 305-326.

Demski, Joel S., and Gerald A. Feltham. “Cost determination: A conceptual approach.” Iowa State University Press (1976).

Friebel, Guido, Matthias Heinz, Miriam Krüger, and Nikolay Zubanov. “Team Incentives and Performance: Evidence from a Retail Chain.” *American Economic Review* 107.8 (2017): 2168-2203.

Holmström, Bengt, and Paul Milgrom. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7 (1991): 24-52.

Jensen, Michael C., and William H. Meckling. “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure.” *Journal of Financial Economics* 3.4 (1976): 305-360.

Katz, Lawrence F., and Alan B. Krueger. “The rise and nature of alternative work arrangements in the United States, 1995-2015” *ILR Review* 72.2 (2019): 382-416.

Lazear, Edward P. “Performance Pay and Productivity.” *American Economic Review* 90.5 (2000): 1346-1361.

Lazear, Edward P. “Compensation and Incentives in the Workplace” *Journal of Economic Perspectives* 32.3 (2018): 195-214.

Levitt, Steven D., and Susanne Neckermann. “What field experiments have and have not taught us about managing workers.” *Oxford Review of Economic Policy* 30.4 (2015): 639-657.

List, John A., and Imran Rasul. "Field experiments in labor economics." *Handbook of Labor Economics*. 4. Elsevier (2011): 103-228.

Manthei, Kathrin, Dirk Sliwka, and Timo Vogelsang. “Performance Pay and Prior Learning: Evidence from a Retail Chain.” *IZA Discussion Paper No. 11859* (2018).

Manthei, Kathrin, Dirk Sliwka, and Timo Vogelsang. “Information Provision and Incentives – A Field Experiment on Facilitating and Influencing Managers’ Decisions.” *Working Paper* (2019).

Mas, Alexandre, and Amanda Pallais. “Labor Supply and the Value of Non-work Time: Experimental Estimates from the Field.” *The American Economic Review: Insights* 1.1 (2019): 111-126.

Prendergast, Canice. “The Provision of Incentives in Firms.” *Journal of Economic Literature* 37.1 (1999): 7-63.

Twenge, Jean M., Stacy M. Campbell, Brian J. Hoffman, and Charles E. Lance. “Generational Differences in Work Values: Leisure and Extrinsic Values Increasing, Social and Intrinsic Values Decreasing.” *Journal of Management* 36.5 (2010): 1117-1142.

Vogelsang, Timo. “Time is (not) Money – Incentive Effects of Granting Leisure Time.” *Working Paper* (2019).

Wiswall, Matthew, and Basit Zafar. “Preference for the Workplace, Investment in Human Capital, and Gender.” *The Quarterly Journal of Economics* 133.1 (2018): 457-507.

Chapter 2

2. Performance Pay and Prior Learning – Evidence from a Retail Chain

2.1 Introduction

Many firms use financial incentives to motivate employees to exert higher efforts (see for instance Prendergast 1999, Lazear 2018 for surveys). Indeed, a still small but increasing number of field studies have shown that performance pay can raise performance significantly in specific environments.[1] However, there is also a substantial share of jobs where performance pay in not used. In his Nobel lecture, Bengt Holmström even states that "*Firms use rather sparingly pay-for-performance schemes.*" (Holmström 2017, p.1769). In the US, for instance, less than 50% of employees work in jobs with performance pay (Lemieux et al. 2009, Bloom and Van Reenen 2011). It is therefore important to advance the understanding for context factors that favor the usefulness of performance pay or limit its benefits.

Studying two field experiments in a retail chain, we identify a limiting factor for the effectiveness of performance pay. We argue that the benefits of introducing performance pay crucially depend on the level of prior learning. In other words, the more experience an organization has formed in a specific stable environment, the smaller the remaining "room for improvement", i.e. potential scope for employees to improve their performance further. As Holmström (2017) has argued, employees are subject to various additional monetary and non-monetary incentives beyond performance pay that influence their behavior. If these forces already constrain employees or drive them to give their best, the opportunity for performance pay to raise performance further may be limited. We formalize the idea that prior learning restricts the benefits of performance pay in a simple model and provide further empirical evidence for this claim.

More precisely, we examine the causal effect of performance pay using two randomized control trials with district (middle-level) managers and later store managers (lower-level) in a German retail firm. The firm operates a large chain of discount supermarkets throughout Germany. Discount supermarkets offer a standard assortment of goods with a strong focus on low prices using standardized processes. The firm employs a store manager for each supermarket, and about

[1] Starting with Lazear (2000) and Shearer (2004), an extensive empirical literature emerged, which is summarized in Bandiera et al. (2011), List and Rasul (2011), Levitt and Neckermann (2015), and Bandiera et al. (2017).

six supermarkets are supervised by a district manager. Hence, there are rather small spans of control and tight central management. Store managers have a limited scope to affect performance but can still acquire knowledge about the specific demand in their store or specific routines that would raise profits. Moreover, their responsibility is to manage the store's workforce, and be accountable for the cleanliness of the stores as well as the presentation of products.

Prior to our study, the central executive management of the chain discussed the usefulness of individual, monetary performance pay in the firm's business model. In collaboration with the regional management, the *average sales per customer ("average receipt")* was identified as a simple and accessible key performance indicator for performance pay in order to generate further incentives to raise the likelihood for a customer to buy more.[2]

In the first experiment, we implemented performance pay based on the average sales per customer for district managers in the fall of 2015.[3] For three months, 25 of 49 randomly selected district managers were eligible to receive this bonus. To filter common exogenous shocks, we used a normalized version of the performance measure relative to each store's own prior development and the development of this key figure in all stores (Holmström 1982, Gibbons and Murphy 1990). Using insights from the first experiment, we implemented the same bonus during the same exact months one year later in 2016 for 194 of 294 store managers. In this second experiment, one treatment replicates the design of the first experiment, and a second treatment uses a simpler bonus formula that reduces the possible complexity of the relative performance evaluation scheme.[4]

We find negligible average treatment effects in both experiments with economically very small upper bounds of 90% confidence intervals (performance increases below 1% or 0.05 standard deviations) in both experiments.

In the spirit of "insider econometrics" (Ichniowski and Shaw 2003), we studied the business in detail, had access to almost all available data from the company, generated survey data through both online surveys with the store managers and telephone interviews with district managers, and continuously analyzed and adjusted the experimental design.

Based on these surveys, we conjecture that store managers' work is characterized by learning about potential improvements (gaining valuable knowledge that increases sales) and habit

[2] The average sales per customer is also known as "average transaction value" or "average customer spent". It is the average sum of sales per customer on a specific visit of a store. For simplicity, we refer to it as the "average sales per customer" in the following.
[3] During the whole experimental period, the company managed the communication (while we prepared everything), and only the senior (top) managers as well as the works council knew that we as researchers were involved. The experiment was called "project," which is a typical wording in the company. In order to control eventual spillovers and avoid potential effects of envy, the control group was also informed that a bonus would be introduced but that the timing of the introduction and the incentivized key performance indicator would vary.
[4] This also relates to the study by Englmaier et al. (2017) in which they changed the communication of a rather complicated incentive scheme and find positive performance effects. While we leave the communication unchanged, we made the incentive scheme easier itself.

formation (acquiring productive routines). We organize this thought in a simple formal dynamic model in which we show that in such an environment, past improvements can limit additional benefits of performance pay. In the model, an agent exerts effort in each period and past efforts increase an agent's future proficiency in doing the job. This naturally leads to concave and bounded learning curves. We then study the effect of introducing performance pay at some later point in time in the learning process and show that the effect of performance pay should be smaller, the later it is introduced. Hence, prior learning limits the added value generated through performance pay and the more efficient a certain process has become, the more difficult it is to generate further performance gains through performance pay.

We explore this idea empirically by studying heterogeneous treatment effects in the second experiment, in which we can access detailed information on prior experience and productivity of stores. To do this, we collect a number of different measures for past experience such as (i) the age of the store, (ii) store managers' tenure, and (iii) age of store managers. We find consistent evidence in line with the hypothesis that performance pay is more effective when there is still "room for improvement". For instance, treatment effects are significantly positive in stores with low levels of experience but become negligible for experienced stores.

With these results, we contribute to the empirical literature on causal effects of financial incentives on employee performance. Using rather simple and repetitive tasks about which an accessible and valid measure of effort is observable, temporary workers, and environments in which monitoring is difficult early studies (some of them non-experimental) find mainly positive effects (see, e.g., Lazear 2000, Shearer 2004, Bandiera et al. 2007, 2009, 2010, Shi 2010). However, as argued by Prendergast (1999) these settings are a good environment for performance pay to work. Banker et al. (2000) provide first evidence that performance pay incentives in the retail sector help to attract and retain high performing employees and increases productivity in general.

Recently, a growing number of studies use firm-level field experiments in more complex environments to study causal effects of incentives on employee performance. In an early study, Casas-Arce and Martinez-Jerez (2009) study a sales contest implemented in a commodities company finding that the introduction increases employees' effort. Delfgaauw et al. (2013, 2014, 2015) run tournament field experiments with a Dutch retailer. Using the average sales growth (Delfgaauw et al. 2013) or the average number of products per customer (Delfgaauw et al. 2015) they find positive performance effects due to implementing a tournament. In contrast to these studies, Delfgaauw et al. (2014) find no average treatment effect when implementing a tournament based on sales revenues in which stores have to outperform comparison stores by a certain amount.

They argue that this is most likely because many leading stores are still behind the required threshold for winning.

Lourenço (2016) analyzes a field experiment in a retail service company and find positive sales increases for individual performance pay. Importantly, this study focuses on sales agents for which the only task is the presentation of products and thus they have a rather simple production function. Friebel et al. (2017) use a field experiment to study the effect of a team bonus in a German bakery chain. They find that the bonus increases sales by 3% which is a third of the sales standard deviation. Interestingly, they show that the treatment effect is larger for stores with a historically larger distance to the target and for stores with a younger workforce. These findings are consistent with our investigation that we only find positive treatment effects of monetary incentives for stores with room for improvement.

We thus add to this literature a firm-level field experiment in a rather difficult environment of the high competitive German discount retailing market and with a complex production function of the incentivized managers. Importantly, we take the important path further of studying individual performance pay in a more complex environment. We show that there exist circumstances under which performance pay does not lead to performance increases and rationalize this using the channel of prior learning and reduced possibilities for further improvements. This finding might explain why in practice we do not see individual performance pay as often as economic theory would suggest.

The literature already acknowledges that performance pay may be less useful in complex work environments. For instance, multitasking distortions can arise because not all aspects of an employee's work are measurable (Holmström and Milgrom 1991, Baker 1992). However, our argument does not rest on the complexity of the environment but rather on its stability; when employees work in stable environments they may build up productive capabilities over time, reducing the value added of performance pay. The paper thus links the literature on performance pay to the literature on human capital formation (e.g. Becker 1962, Becker 1964, Ben-Porath 1967) and learning-by-doing (e.g. Arrow 1962, Jovanovic and Nyarko 1996, Levitt and List 2013). Both strands of the literature argue that knowledge is gradually built up through experience, which leads to concave productivity profiles. Our results can then be interpreted as follows – when past effort leads to more experience and thus knowledge about a stores production function and this builds up persistent human capital, then prior learning can naturally limit the benefits of performance pay.

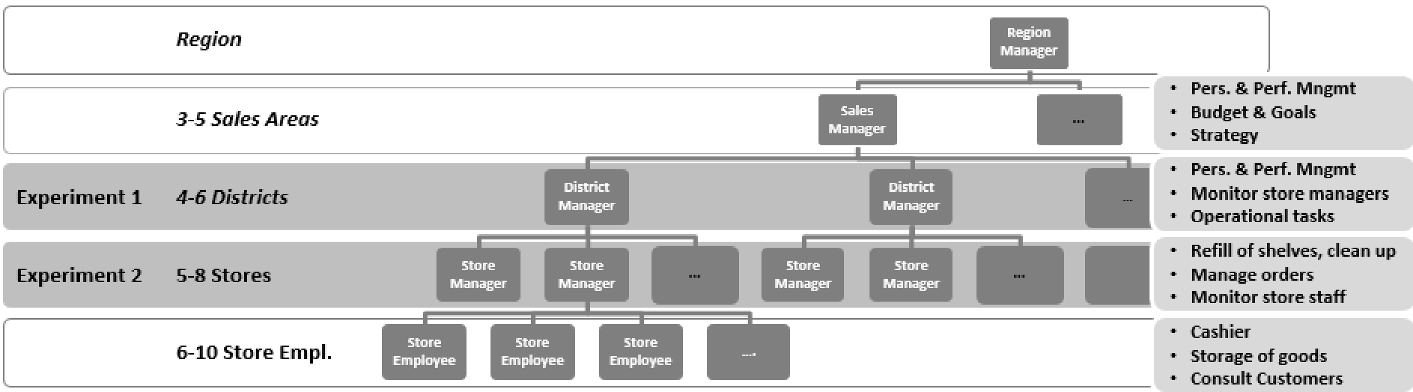
The idea thus closely builds on the role of habit formation in efforts. As documented by Charness and Gneezy (2009) for the case of exercising, monetary incentives can make people develop good habits that persist even when incentives are withdrawn. We argue that the effect also works in the other direction: previously established productive habits may render performance pay dispensable.[5]

The paper proceeds as follows. We first describe the firm and the environment of the field experiments in detail. We then describe the two experimental designs and first key results. Subsequently we develop the formal framework, its implications and go back to the experimental data to study further implications derived from the formal model. The last section concludes.

2.2 The Environment

The company is a large, nationwide retailer operating discount supermarkets in Germany with more than 2,000 stores at the time of the experiment. The supermarket chain is subdivided into several larger geographical regions that cover Germany and has a rather steep hierarchical structure with relatively small spans of control. The structure of the hierarchy is depicted in Figure 2.1. Each region has a regional top manager and is split into sales areas, which are managed by sales area managers. The sales area managers supervise about 4-6 district managers, and the district managers, in turn, are responsible for 5-8 store managers.

Figure 2.1 – Illustrative Organigram of a Sales Region



[5] Our paper is also related on the literature on pay for performance and exploration. Manso (2011) and Ederer and Manso (2013), for instance, have argued that performance pay can reduce incentives to further learn through exploration. Complementary to this, we argue that prior learning also limits the benefits of performance pay.

As is common in discount retailing, the company has highly standardized tasks and processes. Many elements of the store management procedures are determined by the central office (for instance, the store layout and most of the placement of goods).

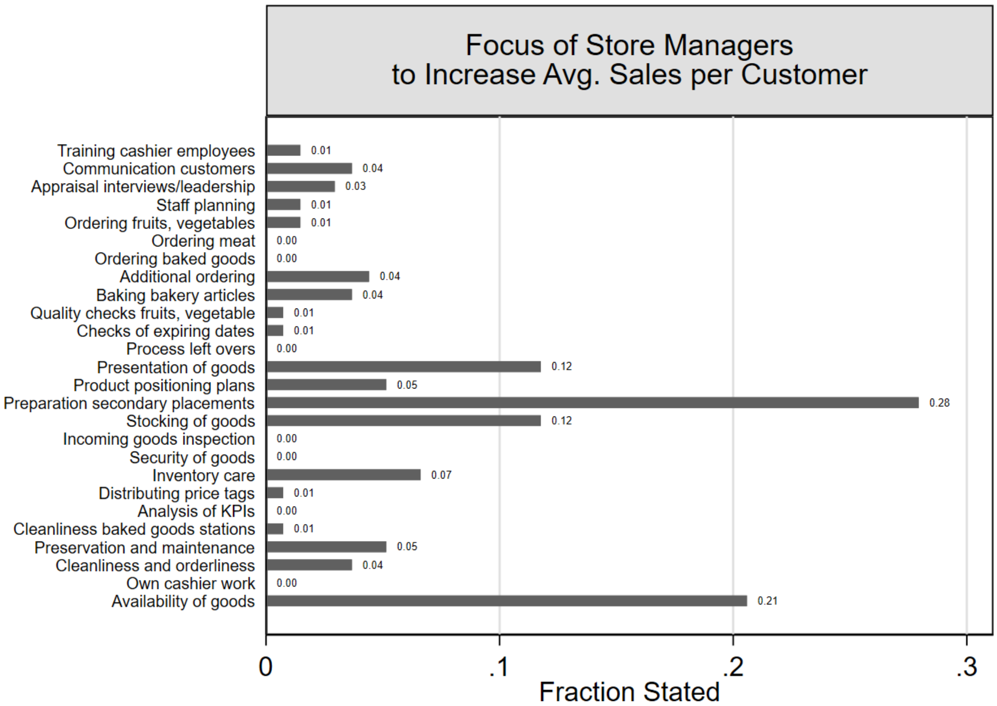
The tasks of sales managers is centered on personnel and performance management. They are also involved in the budget planning procedure and monitoring of relevant KPIs on a regional level to achieve the companies' financial goals. Moreover, they are responsible for implementing the regional strategy and marketing concepts and should ensure that all employees act according to the companies' code of conduct.

District managers are also involved in the personal and performance management of their districts as well as in the budget planning of their district. They generally monitor the store managers but also have some leeway to decide whether to take over operational tasks in the stores or delegate them to store managers. District managers visit their respective store managers approximately twice per week.

Store managers run a store with about 5-8 full time equivalent employees (FTE) and are responsible for the daily operation of the store and execution of operational tasks. This includes guaranteeing that shelves are refilled, the store is kept clean, fresh products (fruits, vegetables and bread) are well presented, and that the cashiers operate efficiently. However, they do have some leeway regarding decisions concerning special placements of goods, temporary price reductions (sales), and product orders where they can overwrite the ordering suggestions made by the computer software using potential local knowledge about customer demand. Moreover, they are involved in the personnel management of the store in cooperation with the district manager and are responsible for the personnel planning. Figure 2.2 shows the possibility of all tasks a store manager could do. It further shows what the store managers did in our experiment to try to increase the relevant KPI of this study (average sales per customer).

Regular store employees are working at the cashier desk, have to store away new incoming goods and should consult customers.

Figure 2.2 - Store Managers' Tasks and Task Focus to Increase Average Sales per Customer



Note: This figure displays all tasks a store managers could possibly execute. It further shows categorized statements from open- questions of an online questionnaire on what store managers did to increase the average receipt of a customer (N=88).

In our meetings with the management prior to the project, we learned that the executive managers had diverse opinions on whether or not monetary incentives could be useful to raise performance in discount retailing. As the firm was considering changing the existing annual bonus scheme for district managers and, more importantly, introducing a bonus scheme for store managers, we proposed to evaluate this question with randomized controlled field experiments. Together with the head office, we approached the regional top manager of one large region with about 300 stores and implemented the two experiments in that region in 2015 and 2016.

2.3 The Experiments

2.3.1 Experiment I: District Managers

2.3.1.1 Design Experiment I

From November 2015 until January 2016, we introduced performance pay by incentivizing an increase in the sales per customer ("average receipt") for a group of randomly assigned district (middle) managers in Western Germany.[6] The district management of this region consisted of 49 managers (covering 300 stores), of which 25 (supervising 152 stores) were randomly assigned to the treatment group using a pairwise randomization method similar to Barrios (2012) and as discussed in Athey and Imbens (2017).[7] The remaining 24 district managers serve as a control group.[8] Table 2.4A in the Appendix shows that randomization was successful with all characteristics not jointly significantly predicting selection into the treatment. In each treatment month, the district managers of the treatment group received €100 (gross for net, approx. 3-5% of their net income) per percentage point increase of the normalized average sales per customer (Norm. Bonus).[9]

From our and the firm's point of view there were several good reasons to incentivize this specific performance measure. First, average sales per customer is a wide-spread figure to measure the success of a store, both for the specific supermarket chain that we studied and in retailing in general (see, e.g., Davids 2013, Bullard 2016). It is often referred to as Average Transaction Value (AVT), Average Dollars per Transaction (ADT), or Average Ticket. It covers the absolute values in quantities sold but also rewards cross-selling. Whenever the number of customers, that enter a particular store, is rather exogenous, the average receipt gives valid indication on the store's

---

[6] As pre-registered at the AEA RCT registry with the ID's AEARCTR-0000961 and AEARCTR-0001758, we also worked with another region for a treatment intervention in which we provide performance feedback without a monetary incentive. However, due to a reallocation of stores to district managers right before the experiment, the treatment and control group are not comparable and empirical estimations with standard models are misleading.

[7] We predicted the average sales per customer for district managers during the treatment period using one year of past data. We then ranked the managers according to this prediction and then randomized treatments within a group of two.

[8] We initially preregistered a sample of 304 stores, but the regional manager removed 4 stores from the pilot (before the start) due to refurbishments and new competitors.

[9] The bonus was a (capped) linear function of the year-on-year percentage point increase in the average sales per customer in the district minus the increase in the average sales per customer of all (more than 2,000) stores in Germany. The district managers received €100 for each percentage point difference above a specific base value, which was equal to the difference of the growth rate of their own district in the first nine months of the year relative to the growth rate of the nation's (Germany) average sales per customer in the first nine months. Thus, both nation-wide shocks and previous performance increases are eliminated. The normalized key figure is:

$$\left(\frac{AvgSalesDistrict_{t,2015}}{AvgSalesDistirct_{t,2014}} - \frac{AvgSalesNation_{t,2015}}{AvgSalesNation_{t,2014}}\right) - \left(\frac{AvgSalesDistrict_{1-9,2015}}{AvgSalesDistirct_{1-9,2014}} - \frac{AvgSalesNation_{1-9,2015}}{AvgSalesNation_{1-9,2014}}\right)$$

As we explain below, we also used a much simpler normalization in our second experiment to address the concern that this might be too complex.

success and allows comparisons between stores of different size and location. Second, it was part of the basic set of key indicators used to steer the company and had been applied throughout company long before we started the experiment. Therefore, it was well-known to the respective managers. The same holds true for the triple normalization that was mapped in the bonus formula. This triple normalization controlled for increases that were already put into effect in the 9 months before the start of the experiment, seasonal variations, as well as for nation-wide shocks in the business cycle. However, rewarding a relative measure might in principle harbor the risk of a ratchet effect as described in Weitzman (1980). Yet, this does not seem likely to occur in our context as the duration and limitation of the 3-months bonus period was transparent to all involved, and hence performance would not affect future targets in any way.

The bonus payment was limited to €500 per month. The bonus for the managers was tripled unexpectedly in the last treatment month (€300 per percentage point increase of the average sales per customer, approx. 10% of their net income), which also lifted the upper cap on payments. No change in the managers' daily business and organizational structure occurred.[10] Managers were not aware that they were taking part in an experiment. During the whole period, we developed the introduction presentation and letters, calculated the bonus, and created monthly notifications. However, in the end company representatives handled all communication of the project. The bonus was introduced during a kick-off meeting with just the managers of the treatment group and communicated again to all district managers by mail.[11]

2.3.1.2 Results Experiment I

In the following, we estimate our main results on the full sample of managers originally assigned to the treatment using a difference-in-difference estimation including fixed effects for months and districts.

$$Y_{dt} = \beta_0 + \beta_1 \cdot Treatment_{dt} + \gamma X_{dt} + a_d + \delta_t + \varepsilon_{dt}$$

[10] District managers had an additional annual bonus plan, which rewarded reduction of inventory losses and personnel expenses. However, this does not conflict with our intervention as it was unchanged and identical for treatment and control group. For the store managers that we study in our second experiment, no such bonus plan existed.

[11] Instructions are provided in the Appendix 2.7.4. Importantly, the managers in the control group knew that other managers received the bonus, but that they would also receive a bonus at some point in the future for a performance variable that was unknown at the time. Possible spillover effects made this communication strategy necessary. The key idea is to avoid managers in the control group feeling unfairly treated upon learning that others receive the bonus. With the bonus being common knowledge, we closely follow Bloom et al. (2015) and Gosnell et al. (forthcoming) and are in line with Bandiera et al. (2011). The company indeed paid out a comparable bonus to the control group in the three months after the end of the treatment.

where Y_{dt} is the average sales per customer in month t for district d. X_{dt} includes time-variant controls which here are dummy variables indicating an ongoing or past refurbishment of the store. ε_{dt} is an idiosyncratic error term clustered at the district level and a_d are district fixed effects.[12] $Treatment_{dt}$ equals 1 for district managers in the treatment group during the treatment period and 0 otherwise. In further specifications we also include district manager and store manager fixed effects. As a baseline specification, we use the time periods from the beginning of the previous year to the end of the experiment (e.g. January 2016 until March 2017, 15 months). Moreover, we provide estimates of the absolute value of the dependent variable. Variations to this are displayed in the Appendix.

Table 2.1 shows results from the fixed effects regressions. As the results show, the treatment had no discernible average effect on performance.[13] Even the upper bound of the 90% confidence interval at €0.056 (approx. 0.44% performance increase; 0.036 standard deviations) is small in terms of economic significance (column 3).[14] Table 2.5A in the Appendix provides robustness checks using ordinary least square regressions (single difference, longer time periods, trimmed data as well as the log of average sales per customer which all confirm this result. [15]

The data of the first two months of the experiment already indicated the main effect to be negligible in size. Therefore, the regional manager decided (upon our request) to triple the amount employees could earn (300€ instead of 100€ per percentage point increase) for the final treatment month (January) to rule out that the incentives were simply too weak to affect behavior (see, e.g., Gneezy and Rustichini 2000). The Appendix shows regression estimates of a monthly regression (Table 2.6A). However, we still find no significant difference between the treatment and the control group in any month and no significant difference between months two and three within the treatment group (Wald test, p = 0.833). Furthermore, Table 2.7A shows no significant treatment effects on any other key outcomes (sales, customer frequency, inventory losses, mystery shopping scores, product ordering behavior, and sick days of store employees). In total, a sum of €5,487.32 was paid out, with an approximate average of €73.16 per district manager per month.

---

[12] We use the allocation of stores to district at the beginning of the experiment as clusters and fix this for the whole estimation period.

[13] Column 3 of Table 2.5A in the Appendix displays results from a regression with trimmed data (top and bottom 1%) and shows that the negative sign of the coefficient might depend on some outliers in the data.

[14] As ex-post power calculations to support null effects are problematic (Hoening and Heisey 2001), we prefer to refer to the confidence intervals to illustrate the possible range of effects (see, e.g., Groth et al. 2016).

[15] Note that this effect is very small also in comparison to the effects of performance pay reported in the literature so far. For instance, Friebel et al. (2017) estimate an effect of a team bonus in a bakery chain of 0.3 standard deviations and Bandiera et al. (2017) estimate an average effect of performance pay of 0.28 standard deviations using a meta-analysis.

Table 2.1 – Main Effects Experiment I&II

	Experiment I – District Level			Experiment II – Store Level		
	(1)	(2)	(3)	(4)	(5)	(6)
	Sales per Customer	Sales per Customer	CI 90%	Sales per Customer	Sales per Customer	CI 90%
Treatment Effect *Norm. Bonus*	0.0020 (0.0464)	-0.0240 (0.0475)	[-0.1037; 0.0556]	-0.0162 (0.0437)	-0.0099 (0.0478)	[-0.0902; 0.0703]
Treatment Effect *Simple Bonus*				0.0328 (0.0504)	0.0347 (0.0594)	[-0.0649; 0.1343]
Time FE	Yes	Yes		Yes	Yes	
Store/District FE	Yes	Yes		Yes	Yes	
District Manager FE	No	Yes		No	Yes	
Store Manager FE	No	No		No	Yes	
N of Observations	637	637		3822	3473	
Level of Observations	District	District		Store	Store	
N of Districts/ Stores	49	49		294	294	
Cluster	49	49		50	50	
Within R^2	0.9427	0.9478		0.8473	0.8476	
Overall R^2	0.1043	0.1185		0.0497	0.0327	

Note: The table reports results from a fixed effects regression with the sales per customer on the district/ store level as the dependent variable. The regression accounts for time and store district fixed effects and adds fixed effects for district managers in column 2 and fixed effects for district and store managers in column 5. For experiment I, the regressions compare pre-treatment observations (January 2015 - October 2015) with the observations during the experiment (November 2015 – January 2016). For experiment II, the regressions compare pre-treatment observations (January 2016 - October 2016) with the observations during the experiment (November 2016 – January 2017). *Treatment Effect* thus refers to the difference-in-difference estimator. All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. Columns 3 and 6 display 90% confidence intervals of the specification in column 3 and 6, respectively. * p<0.1, ** p<0.05, *** p<0.01.

2.3.1.3 Post-Experimental Interviews

To investigate possible reasons for the absence of a meaningful treatment effect, we conducted a telephone survey in June 2016 and interviewed 19 of the 25 treated district managers on behalf of the company.[16] All district managers reported having tried to influence the average sales per customer. Still district managers claimed that it is necessary to delegate the tasks to store managers to influence the average sales per customer. Hence, it is conceivable that the bonus would be more effective when targeted at the store managers, who are more immediately responsible for operating the stores.[17]

[16] Of the 25 district managers in the first period, 3 have left the company and 3 refused to talk to us unless they had formal written permission from the regional manager.

[17] Indeed, the post-experimental questionnaire of Experiment II confirmed that store managers themselves state that they have more influence on the average sales per customer than district managers.

2.3.2 Experiment II: Store Managers

2.3.2.1 Design Experiments II

Based on the above insight, we ran a follow-up experiment one year later in the same calendar months (November 2016 – January 2017), now incentivizing store managers. We held the circumstances constant and used the same performance measure – only this time measured at the store level. We now compare a control group to two different treatment groups: One treatment group received a bonus based on exactly the same formula as before (*Norm. Bonus*) but applied for the store managers, whereas the other one was subject to a substantially simpler year-on-year comparison (*Simple Bonus*).[18] The key idea of the second treatment was to investigate whether the normalization led to an overly complex bonus formula, which may have limited its impact on performance.[19]

We used the same pairwise randomization method as in Experiment I to create new treatment groups and randomly assign stores within districts. This leads to 95 stores in the group with the bonus calculation method used previously for the district managers (*Norm. Bonus*), 95 stores in the group with the simplified year-on-year calculation (*Simple Bonus*) and 99 stores in the control group. The balancing table (Table 2.8A) shows the successful randomization.

Each month, store managers received €125 (approx. 4% of their gross income) per point increase of the respective normalized average sales per customer.[20] The bonus payment was limited to €375 per month.[21] As before, all communication was standardized and handled by company representatives. We used the same communication strategy, material and wording as in Experiment I.[22]

---

[18] The simplified key figure here is: $\frac{AvgSalesStore_{t,2016}}{AvgSalesStore_{t,2015}}$

[19] On the other hand, a preregistered countervailing effect could be that managers positively reciprocate the normalized bonus because they feel better insured.

[20] The difference to Experiment I occurs because this time taxes had to be paid on the bonuses, but the relation to the monthly salary is similar. The reason for the net bonus in the case of district managers was that the company could use a tax exemption – the transfer was made through a company shopping card – which was not feasible for store managers.

[21] In contrast to the district managers, store managers were previously not eligible for any bonus.

[22] The only difference is that this time the communication was done by letters sent through the standard postal service as emails to store managers could be accessed by all store employees. Additionally, we received the full support of the works council.

2.3.2.2 Results Experiments II

Again, Table 2.1 shows results from a fixed effects regression, with the store level being the unit of observation.[23] Column 4 shows a point estimate without controlling for district manager and store manager fixed effects. Column 5 controls for district managers and store manager fixed effects. Again, the effects of both treatments are not only statistically insignificantly different from 0 (and from each other) but also economically very small. As before, upper bounds of the 90% confidence intervals are economically very small at approx. 1% (0.0545 standard deviations) and 0.5% (0.0285 standard deviations), respectively. Robustness checks are again displayed in the Appendix (Table 2.9A), monthly treatment effects in Table 2.10A. Table 2.11A shows possible influences on other key outcomes (sales, customer frequency, inventory losses, mystery shopping scores, product ordering behavior, and sick days of store employees) with no significant treatment effect. In total, a sum of €68,221.98 was paid out as bonus payments, with an average of approximately €108.39 per store manager per month.

2.3.2.3 Post-Experimental Survey and Interviews

At the end of the second experiment (end of January), we invited all store managers to participate in an online survey.[24] In total 43.20% of all store managers answered all questions of the survey.

Concerning satisfaction with work, salary, work stress, employer fairness, and life in general, we do not find any statistical significance difference between the three groups of experiment II. Therefore, it seems unlikely to have negative influences on the managers. As already mentioned above, managers from all groups stated that the average sales per customer can be more easily influenced by store managers than by district manager (p<0.001).

Comparing the two respective bonus schemes, there are statistically significant differences in store managers' perceptions of the respective scheme (Appendix Table 2.12A).[25] Most importantly, store managers perceived the normalized bonus formula as more complicated

[23] At the request of the company and to be consistent with Experiment I, we only assigned the treatment to stores older than two years, which lead to a reduction of the treated stores from the preregistered sample. Accidentally, two younger stores were assigned a treatment, but this was corrected by the company afterwards. As before, we only include stores in the regression that have been open for more than two years in order to make all three groups comparable. Data for store managers who switch stores during the treatment period are dropped from the analysis. Including the full sample does not lead to qualitative differences in the results.

[24] This was the first time we became apparent as a university as we officially conducted the surveys to maintain anonymity of the managers.

[25] Store managers in both bonus treatments had to respond to the same survey items containing statements about the bonus formula such as "The bonus formula was fair", "I understood the bonus formula" or "The bonus formula was complicated". Store managers had to evaluate the statements on a scale from 1 = completely agree to 6 = completely disagree.

(p<0.01) and not easy to understand (p<0.01). Interestingly, store managers in the treatment with the normalized bonus formula perceived the bonus formula to be as fair as those in the treatment with the simple bonus (p<0.01). Importantly, they generally agree that they know how to influence average sales per customer (Wilcoxon Signed-Rank test against a neutral response of 3, p<0.001).

We also included open-ended questions in the online survey with the store managers and in January and February 2017, we again conducted telephone interviews with all district managers. After the end of the treatment intervention, we asked store managers in open-ended survey questions for potential difficulties in influencing the average sales per customer. Exemplary statements of store managers are:

- "*No leeway. Strict predetermined concept.*"
- "*The given placements by the district manager. The store managers know better what sells well.*"
- "*I do my best every day and thus a further increase was simply impossible*."
- *"A high average receipt from the beginning [...]."*
- *"High average receipt, low customer frequency."*
- *"Because in my store all shelves are always filled, I couldn't do more."*
- *"Not a lot of room for my own ideas."*
- *"I already have a high average receipt and due to [competitor X] also less sales."*

Exemplary statements in the interviews with the district managers after the end of Experiment II are:

- *"A high average receipt from the start [...]."*
- *"If the store manager already did a good job and implemented all things, then the store manager has a high average receipt and a further increase is difficult as the leeway is restricted."*
- *"The store managers will be incentivized, but it is extremely difficult to raise the average receipt if it's already on a high level."*
- *"[...] Store manager did a good job throughout the whole year to increase the average receipt, but it is simply not possible for him to raise it further in the required months."*
- *"My store managers have been trying to increase the average receipt for years with great success. Now it is much more difficult to perform during the bonus period."*

Hence, the main aspects that managers mentioned were limited autonomy, their own activities prior to the introduction of the bonus, and past efforts that had been invested to raise the average sales per customer that leave little further potential.

2.4 Prior Learning and Performance Pay

2.4.1 A Conceptual Framework

A key argument that is repeatedly mentioned by managers' in the survey is that in their limited scope to raise the average sales per customer, they have already put numerous measures into practice before. Therefore, it was claimed that the respective potential to improve further tended to be exhausted. The environment thus seems to be characterized by a combination of "learning-by-doing" (Arrow 1962, Jovanovic and Nyarko 1996, Levitt and List 2013) and habit formation in efforts (Charness and Gneezy 2009). Intuitively, store managers learn over time how to raise the average sales per customer and establish routines that carry over into future periods.

We now explore a simple model to illustrate this idea and its implications. The performance of an organizational unit in period t is a function of the agent's proficiency p_t in managing the unit. Profits in period t are given by

$$f(p_t)$$

where $f'(p) > 0$ and $f''(t) \leq 0$. In each period the agent can exert an effort e_t at cost $c(e_t)$ where $c''(e_t) > 0$ and $c'(\overline{e}) = 0$ for some $\overline{e} > 0$.[26] The agent's proficiency in period t is a function of her prior proficiency p_{t-1} and the effort exerted in the current period t

$$p_t = \phi p_{t-1} + \gamma e_t$$

with $0 < \phi, \gamma < 1$. Hence, efforts exerted in a given period raise performance in that period but also may generate more persistent effects on future performance. The parameter γ measures the marginal returns to current efforts and ϕ captures the level of habit formation or human capital acquisition. When ϕ is larger, efforts form habits to a stronger extent.[27] If, for instance, $\phi = 0$, the

[26] Hence, the agent's cost function is first decreasing and then increasing in effort. We thus assume that the agent voluntarily exerts some effort even in the absence of any formal incentives (for instance because she may to some extent be intrinsically motivated or because of monitoring and firing threats).

[27] Note that the model can be equivalently transformed to one in which the agent chooses k_t at costs $c\left(\frac{k_t - \phi k_{t-1}}{\gamma}\right)$ which is close to common representations of habit formation in consumer theory and macroeconomics (see, e.g. Ravn et al. 2006).

model is a standard moral hazard model with purely transitory efforts. If $\phi = 1$, then efforts are fully persistent human capital investments. If $0 < \phi < 1$ then efforts are habit forming or generate human capital, but there is human capital depreciation, i.e. agents forget knowledge or partially lose habits or routines when not investing further efforts.

We first analyze the dynamics of store performance when there is no performance pay. In this case, the agent exerts effort $e_t = \overline{e}$ in each period. Hence,

$$p_t = \gamma\overline{e}\sum_{\tau=0}^{t-1}\phi^{\tau}$$

which corresponds to the sum of a finite geometric series such that

$$p_t = \gamma\overline{e}\frac{1-\phi^t}{1-\phi}.$$

Hence, we obtain the following result:

Proposition 1: *When there is no performance pay, profits in period t are given by*

$$f\left(\gamma\overline{e}\frac{1-\phi^t}{1-\phi}\right).$$

Profits are increasing over time and converge to $f\left(\frac{\gamma\overline{e}}{1-\phi}\right)$.

The simple model thus implies an increasing and bounded learning curve. In each period the agent exerts some effort and learns from experience.

Now suppose that a bonus β is introduced in period t for one period. The agent now maximizes

$$\max_{e_t} \beta f(\phi p_{t-1} + \gamma e_t) - c(e_t)$$

with first order condition

$$\beta f'(\phi p_{t-1} + \gamma e_t)\gamma - c'(e_t) = 0$$

which implicitly defines effort in period t as a function of the bonus and prior knowledge $e_t(\beta, p_{t-1}, \gamma)$. This leads to the following result:

Proposition 2: *When there are decreasing returns to proficiency (i.e. $f''(p_t) < 0$), the performance effect of introducing a bonus in period t will be decreasing in t.*

Proof:

The performance gain from incentives is equal to

$$\Delta\pi = f\big(\phi p_{t-1} + \gamma e_t(\beta, p_{t-1}, \gamma)\big) - f(\phi p_{t-1} + \gamma\overline{e})$$

and

$$\begin{aligned}\frac{\partial\Delta\pi}{\partial p_{t-1}} &= f'\big(\phi p_{t-1} + \gamma e_t(\beta, p_{t-1}, \gamma)\big)\left(\phi + \gamma\frac{\partial e_t(\beta, p_{t-1}, \gamma)}{\partial p_{t-1}}\right) - f'(\phi p_{t-1} + \gamma\overline{e})\phi \\ &= \Big(f'\big(\phi p_{t-1} + \gamma e_t(\beta, p_{t-1}, \gamma)\big) - f'(\phi p_{t-1} + \gamma\overline{e})\Big)\phi \\ &\quad + f'\big(\phi p_{t-1} + \gamma e_t(\beta, p_{t-1}, \gamma)\big)\gamma\frac{\partial e_t(\beta, p_{t-1}, \gamma)}{\partial p_{t-1}} \qquad < \ 0\end{aligned}$$

as by the implicit function theorem

$$\frac{\partial e_t}{\partial p_{t-1}} = -\frac{\beta f''(\phi p_{t-1} + \gamma e_t)\gamma}{\beta f''(\phi p_{t-1} + \gamma e_t)\gamma^2 - c''(e_t)}\phi < 0.$$

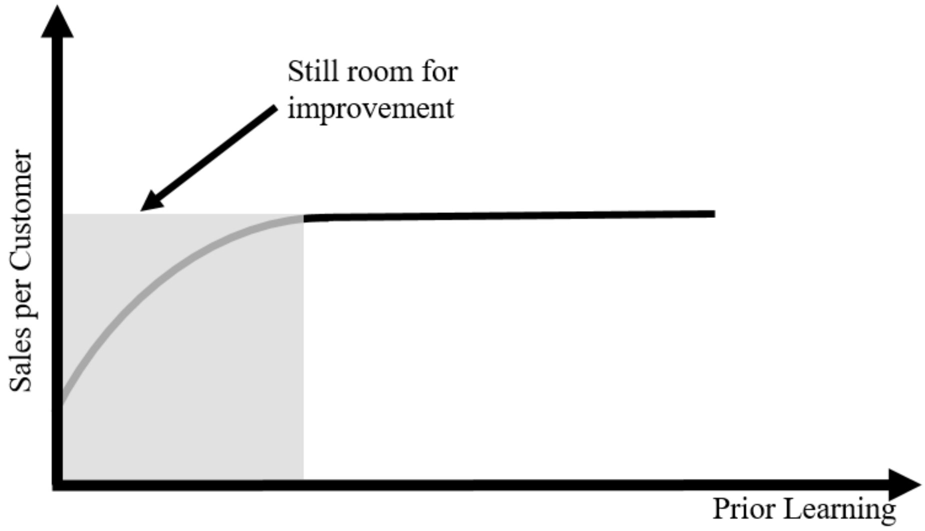
As p_{t-1} is increasing in t the result follows. ■

When there is learning-by-doing or habit formation, performance pay thus has a stronger effect on performance when agents are still early on in the learning curve. The more knowledge, routines, or productive habits an agent has acquired before, the weaker the additional gain from exerting more effort. When $f(p_t)$ is bounded (for instance if agents have limited job scope), then $\lim_{p_{t-1}\to\infty}\Delta\pi = 0$ such that performance pay can become ineffective for agents with strong experience. We explore these implications empirically in the next section.

2.4.2 Empirical Evidence

A straightforward conjecture based on the model is thus that the bonus had negligible effects because earlier activities reduced the scope to increase the sales per customer further. However, if this is indeed the case, we should be able to detect an effect of the bonus, for those stores that are "early on" in the learning curve. The key idea is illustrated in Figure 2.3. The closer a manager is to the beginning of the learning curve (less prior learning), the more room for improvement exists.

Figure 2.3 – Illustration Learning Curve



A first simple implication of the model is that store managers should find it harder to increase average sales per customer when average sales per customer are higher. This idea is supported by the questionnaire data reported in Table 2.12A in the Appendix. In each of the three treatment groups store managers state that it is easier to influence the average sales per customer with initially low rather than initially high average sales per customer ($p<0.01$).[28]

In a next step, we now explore the hypothesis that (i) treatment effects are positive for stores with a low experience and that (ii) treatment effects decrease with experience. The empirical model we estimate to investigate these heterogeneous treatment effects is the same fixed effects

[28] To be precise: The respective survey items are "A store with an initially high average receipt can more easily influence the average receipt." and "A store with an initially low average receipt can more easily influence the average receipt". In all three groups store managers agree significantly more often to the second item.

difference-in-difference regression as before. Only this time we additionally interact the treatment variable with proxies for prior experience.

$$Y_{st} = \beta_0 + \beta_1\, Treatment_{st} + \beta_2 Treatment_{st} \times Experience_s + \gamma X_{st} + \delta_t + \delta_t \times Experience_s + \delta_s + \delta_b + \varepsilon_{s,t}$$

To allow for different time trends of stores of different levels of experience we also include interaction terms of the experience proxies with the time fixed-effects. We apply different normalizations of experience to investigate not only the heterogeneous treatment β_2, but to study the size of the treatment dummy β_1 in stores with low experience. We estimate this for both performance pay treatments separately (i.e. both bonus formulas that were implemented in the second experiment).

We measure experience by (1) the age of the store, (2) the tenure of the store manager in the firm, and (3) the age of the manager. We compute the percentile value (the value of the cumulative distribution function) of each of these variables[29] and start by interacting the treatment with the average experience percentile (i.e. the mean of the percentiles of age of the store, tenure of the manager, and age of the manager).

The regression results are reported in Table 2.2. In line with the conjecture that the bonus is less effective later on in the learning curve, the interaction terms are significantly negative in both treatments. Hence, the size of the treatment effect is decreasing with experience. Note that the treatment coefficients estimate the effect of the treatment in a store which would have the lowest experience in all three proxy variables. The estimate amounts to an increase in sales per customer of about €0.32 or about 2.4% (p<0.02, Table 2.2, Column 2) in both treatment groups.

Table 2.13A in the Appendix reports robustness checks (single difference, longer time periods, trimmed data, log values) and Table 2.14A displays a regression where we interact each experience proxy separately in the regression.[30]

[29] To be precise: The respective variable is the rank of the store with respect to the proxy (starting with the store with least experience) divided by the number of all stores such that the variable takes value 1 for the store with the highest experience and takes a value close to zero for the stores with the lowest experience. See, for instance, Aggarwal and Samwick (1999) for a similar approach.

[30] Note that there is no statistically significant correlation between the three proxies that cover personal and store characteristics (Spearman rho between *Age Manager* and *Age Store* = 0.0477, p = 0.4132, Spearman rho between *Tenure Manager* and *Age Store* = 0.0272, p = 0.6430). But store manager age and tenure are of course positively correlated (Spearman rho between *Tenure Manager* and *Age Manager* = 0.5295, p <0.001).

Table 2.2 – Heterogeneous Treatment Effects by Experience

	Sales per Customer	
	(1)	(2)
Treatment Effect Norm. Bonus	0.270^{**} (0.122)	0.324^{**} (0.134)
Treatment Effect Norm. Bonus x *Experience Proxy*	-0.539^{**} (0.206)	-0.632^{***} (0.233)
Treatment Effect Simple Bonus	0.260^{**} (0.122)	0.338^{**} (0.131)
Treatment Effect Simple Bonus x *Experience Proxy*	-0.435^{**} (0.212)	-0.578^{**} (0.235)
Time FE x *Experience Proxy*	Yes	Yes
Time FE	Yes	Yes
Refurbishments	Yes	Yes
Store FE	Yes	Yes
District Manager FE	No	Yes
Store Manager FE	No	Yes
N of Observations	3692	3378
N of Stores	284	284
Within R^2	0.8474	0.8486
Overall R^2	0.0514	0.0359

Note: The table reports results from a fixed effects regression with sales per customer on the store level as the dependent variable. The regression accounts for time and store fixed effects in column 1 and adds district manager and store manager fixed effects in column 2. The regressions compare pre-treatment observations (January 2016 - October 2016) with the observation during the experiment *TreatmentTime* (November 2016 – January 2017). All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. *Experience Proxy* (between 0 and 1) refers to the mean percentile of a store's age, manager's tenure, and manager's age of the respective manager/store. The regressions interact all time variables with the *Experience Proxy*. Note that for 10 observations we do not have date on job tenure. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Note that the main treatment effect is here estimated for a (hypothetical store) at the lowest end of the experience distribution and that this estimation hinges on the assumption that the interaction effect is linear in experience. It is therefore important to check the robustness of the results when we investigate treatment effects directly for subsamples of stores with low experience. We estimate the treatment effects separately within the group of stores where the mean percentile of the experience proxies is below 30%, 40%, 50%, and 60%, respectively. Table 2.3 reports the respective regressions of average sales per customer on treatment dummies in the different subsamples. As column (1) shows, both treatments have sizeable (>€0.30) and highly significant (p<0.01) effects in the group of stores where the mean percentile of the experience proxies is below 30%. The effect is still significant for stores where the mean percentile is below 50% but then has only about half the magnitude.

Table 2.3 – Treatment Effects in Stores With Low Experience

	Cut-Offs of the Experience Proxy			
	(1)	(2)	(3)	(4)
	<=0.3	<=0.4	<=0.5	<=0.6
Treatment Effect	0.309***	0.198**	0.166**	0.0237
Norm. Bonus	(0.110)	(0.0933)	(0.0688)	(0.0642)
Treatment Effect	0.369***	0.168*	0.176**	0.0786
Simple Bonus	(0.119)	(0.0868)	(0.0693)	(0.0664)
Time FE	Yes	Yes	Yes	Yes
Refurbishments	Yes	Yes	Yes	Yes
District Manager FE	Yes	Yes	Yes	Yes
Store Manager FE	Yes	Yes	Yes	Yes
N of Observations	521	1128	1748	2222
N of Stores	45	96	148	189
Within R^2	0.8840	0.8824	0.8631	0.8573
Overall *R2*	0.0686	0.0846	0.0468	0.0225

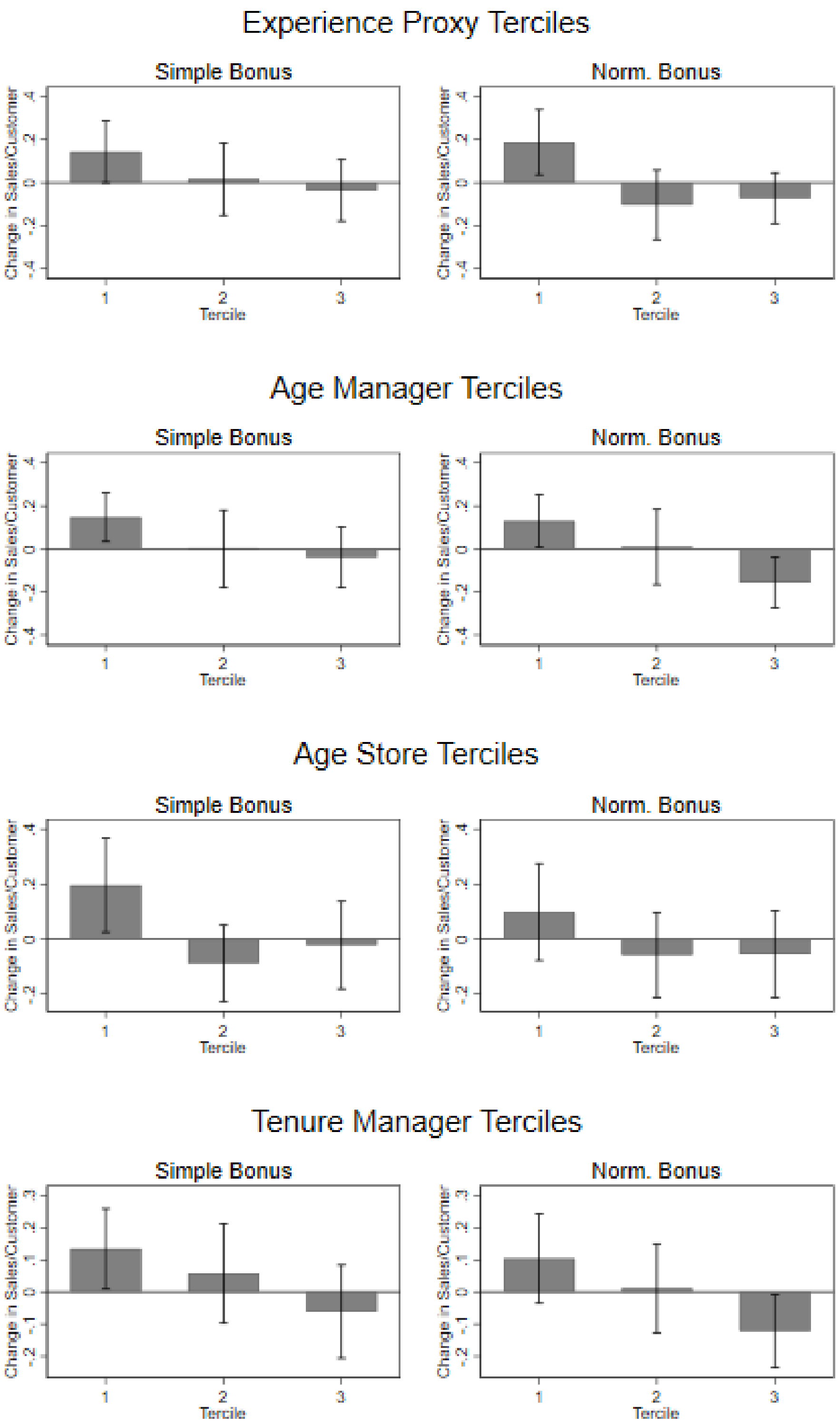
Note: The table reports results from a fixed effects regression with sales per customer on the store level as the dependent variable in different subsamples of the *Experience Proxy*. *Experience Proxy* refers to the mean percentile of a store's age, manager's tenure, and manager's age of the respective manager/store. The regression accounts for time, district, district manager, and store manager fixed effects. The regressions compare pre-treatment observations (January 2016 - October 2016) with the observation during the experiment *TreatmentTime* (November 2016 – January 2017). All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. We start at <=0.3 because we only have 13 stores with <=0.2. * p<0.1, ** p<0.05, *** p<0.01.

Finally, for all four indicators that we used (mean percentile of experience proxies, and age manager, tenure manage, and age store) we estimated treatment effects in each tercile of the distribution of the respective experience measure. These estimates are displayed in Figure4. For each of the four indicators and two treatments, the point estimates are largest in the lowest tercile and are smaller for higher values of the respective proxy.

As the Figure 2.4 shows, the effect of the simple bonus essentially becomes zero in the largest experience terciles. It also indicates that the normalized bonus may even have had a negative effect in stores with high experience. A potential explanation for this observation is the following: In this treatment, store managers earned a bonus only when exceeding a threshold of sales per customer determined directly before the intervention. Hence, this scheme made it particularly hard for store managers who had been successful in raising the key figure already before the intervention. It is conceivable that this induced a demotivating effect as store managers may have felt punished for past successes.[31]

[31] Recall that store managers who received the normalized bonus considered the bonus significantly less fair than those who received the simple bonus (see section 2.3.2.3).

Figure 2.4 - Treatment Effects by Terciles of Experience Proxies



Note: This figure displays treatment effects on sales per customer for different experience variables in different terciles with 90% confident intervals. To estimate treatment effects, we generate dummies for the different treatments and the different terciles of the experience variable and regress sales per customer on these dummies using a fixed effects regression with time, store, district manager and store manager fixed effects. The regression compares pre-treatment observations (January 2016 - October 2016) with the observation during the experiment *TreatmentTime* (November 2016 – January 2017). The regression controls for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period.

2.5 Conclusion

We report two firm-level field experiments in a retail chain showing that individual performance pay may not always raise performance in an economically meaningful way. We did not find a positive average treatment effect of the performance-contingent bonus on the incentivized key figure (sales per customer) for district managers. We then replicated this finding for store managers. Results from surveys and interviews suggest that past activities already had raised sales per customer to a level that had made it hard for store managers to achieve further increases. We rationalized this conjecture in a framework in which prior learning and habit formation can generate persistent effects of effort on performance. As we show, in such a framework prior learning can naturally limit the performance effects of performance pay. We then explored implications of the model in further analyses of the data from the field experiments. Most importantly, we find that performance pay raised performance in stores with little prior experience (i.e. young stores with young store managers) but that treatment effects vanish with experience.

Our results thus point to a further explanation that contributes to our understanding for the absence of performance pay in many jobs beyond the typically stated multitasking distortions or a lack of available performance measures: Even if there are no such distortions and clean and simple performance measures are available, prior learning and the formation of productive habits or routines may in stable environments leave little room to raise performance further. Bonus payments can, however, lead to performance increases in areas where room for improvement (still) exists.

We do not claim that our results are more representative for the question of whether performance pay raises performance than previous field experiments, but we assert that they are not less representative. In other words, we view the results as a cautionary tale. Performance is often driven (or constrained) by many other management practices, company policies and regulations, or social norms of behavior. In some cases, performance pay may not be able to affect performance to a significant extent beyond the already achieved.

A further implication is that, in order to extrapolate the effects of performance pay as estimated in a specific study, it is important to take the prior experience of the respective workforce into account. In lab experiments or in field experiments conducted with temporary workers, for instance, subjects typically face novel tasks where learning curves can be steep. Hence, these studies should rather yield upper bounds for the performance effects than what could be expected among more experienced workers. It even seems conceivable that the large performance effects of about 20% identified in Lazear (2000) are to some extent due to Safelite's rather inexperienced workforce. Safelite's turnover rates were over 4.5 percent per month and the average tenure of the workforce was only about two-thirds of a year (Lazear 2000, p. 1354).[32] As our model suggests, such an environment should be a particularly fertile ground for strong performance effects of bonus payments.

Our results also have broader implications for the design of bonus schemes in practice. At first is seems to be important that companies use key figures in their bonus schemes or for performance evaluation in general that can actually be influenced by managers. With unchanged technology and managers who are continuously trying to improve, key figures have an upper bound at which it is useless to incentivize them further. However, what seems straightforward at the first sight might be very difficult to investigate in practice.

Moreover, given our results it seems to be reasonable to incentivize employees at the early stage of their career. If incentives increase learning of employees leading to a persistent higher performance, this can help to increase knowledge about the underlying production functions for unexperienced employees. An open question is still whether one should remove the incentives again when managers are experienced enough. The theoretical idea presented above suggests that removing the incentive should leave performance on the same level due to persistent learning and performance increases. However, different psychological mechanisms might also play a role than and could lead, for instance, to crowding out effects.

In a similar vein, our results can then help to understand why firms quite frequently change incentive schemes or the underlying key figures used to measure performance.[33] As mentioned above, standard principal agent models suggest that in stable environments there is an optimal set of key figures that should be used for incentive compensation as long as the underlying technology does not change. But if there are bounded learning curves and agents keep up acquired productive habits and routines, it may become beneficial to vary the performance indicators used in incentive compensation over time in order to focus employee's attention to form new habits on routines for tasks where there is still room for further improvement.

[32] As Lazear and Shaw (2008, p. 708) document, workers at Safelite faced steep learning curves and workers at their first month of tenure were 42% less productive than the same workers one year later.
[33] For their higher-level managers, the firm we study for instance changed the key figures used for incentive compensation every year.

2.6 References of Chapter 2

Aggarwal, Rajesh K. and Andrew A. Samwick. "The Other Side of the Trade-off: The Impact of Risk on Executive Compensation." Journal of Political Economy 107.1 (1999): 65-105.

Arrow, Kenneth J. "The Economic Implications of Learning by Doing". Review of Economic Studies 29.3 (1962): 155-173.

Athey, Susan and Guido Imbens. "The Econometrics of Randomized Experiments." in A. Benerjee and E.Duflo. Handbook of Field Experiments. 1. Elsevier (2017): 73-140

Baker, George P. "Incentive Contracts and Performance Measurement." Journal of Political Economy 100.3 (1992): 598-614.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Incentives for managers and inequality among workers: evidence from a firm-level experiment." Quarterly Journal of Economics 122.2 (2007): 729-773.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Social Connections and Incentives in the Workplace: Evidence From Personnel Data." *Econometrica* 77.4 (2009): 1047-1094.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Social Incentives in the Workplace." *The Review of Economic Studies* 77.2 (2010): 412-458.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Field experiments with firms." *Journal of Economic Perspectives* 25.3 (2011): 63-82.

Bandiera, Oriana, Greg Fischer, Andrea Prat, and Erina Ytsma. "Do Women Respond Less to Performance Pay? Building Evidence from Multiple Experiments." *Working Paper* (2017).

Banker, Rajiv D., Seok-Young Lee, Gordon Potter, and Dhinu Srinivasan. "An Empirical Analysis of Continuing Improvements Following the Implementation of a Performance-Based Compensation Plan." *Journal of Accounting and Economics* 30.3 (2000): 315-350.

Barrios, Thomas. "Optimal Stratification in Matched Pairs." *Working Paper* (2014).

Becker, Gary. "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy* 70.5 (1962): 9-49.

Becker, Gary. "Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education." University of Chicago Press (1964).

Ben-Porath, Yoram. "The production of human capital and the life cycle of earnings." *Journal of Political Economy* 75.4 (1967): 352-365.

Bloom, Nicholas, James Liang, John Roberts, and Zhichung Jenny Ying. "Does Working from Home Work? Evidence from a Chinese Experiment." *Quarterly Journal of Economics* 130.1 (2015): 165-217.

Bloom, Nicholas and John Van Reenen. "Human Resource Management and Productivity." *Handbook of Labor Economics* (2011): 1697-1767.

Bullard, Brittany. “Style and Statistics: The Art of Retail Analytics”. Wiley and SAS Business Series (2016).

Casas-Arce, Pablo, F. Asis Martinez-Jerez. “Relative Performance Compensation, Contests, and Dynamic Incentives.” *Management Science* 55.8 (2009): 1306-1320.

Charness, Gary, and Uri Gneezy. “Incentives to Exercise.” *Econometrica* 77.7 (2009): 909-931.

Davids, John A. “Measuring Marketing: 110+ Key Metrics Every Marketer Needs.” Wiley (2013).

Delfgaauw, Josse, Rober Dur, Arjan Non, and Verbeke, Willem. "Dynamic Incentive Effects of Relative Performance Pay: A Field Experiment." *Labor Economics* 28.1 (2014): 1-13.

Delfgaauw, Josse, Rober Dur, Arjan Non, and Verbeke, Willem. "The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment." *Journal of Labor Economics* 33.3 (2015): 521-569.

Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke. "Tournament incentives in the field: Gender differences in the workplace." *Journal of Labor Economics* 31.2 (2013): 305-326.

Ederer, Florian, and Gustavo Manso. “Is Pay-for-Performance Detrimental to Innovation?” *Management Science* 56.7 (2013): 1496-1513.

Englmaier, Florian, Andreas Roider and Uwe Sunde. “The Role of Communication of Performance Schemes: Evidence from a Field Experiment.” *Management Science* 63.12 (2017): 3999-4446.

Friebel, Guido, Matthias Heinz, and Nick Zubanov. "Team incentives and performance: Evidence from a retail chain." *American Economic Review* 107.8 (2017): 2168-2203.

Gibbons, Robert, and Murphy, Kevin J. “Relative performance evaluation for chief executive officers.” *Industrial and Labor Relations Review* 43.3 (1990): 30-S.

Gibbs, Michael, Susanne Neckermann, and Christoph Siemroth. “A Field Experiment in Motivating Employee Ideas” *Review of Economics and Statistics* (2017).

Gneezy, Uri, and Aldo Rustichini. "Pay enough or don't pay at all." *Quarterly Journal of Economics* 115.3 (2000): 791-810.

Gosnell, Greer K., John A. List, and Robert Metcalfe. “The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains.” *Journal of Political Economy* (forthcoming).

Groth, Matthew, Nandini Krishnan, David McKenzie, and Tata Vishwanath. “Do Wage Subsidies Provide a Stepping-Stone for Recent College Graduates? Evidence from a Randomized Experiment in Jordan.” *Review of Economics and Statistics* 98.3 (2016): 488-502.

Hoenig, John M., and Dennis M. Heisey. “The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis.” *The American Statistician* 55.1. (2001): 19-24.

Holmström, Bengt. “Moral hazard in teams.” *Bell Journal of Economics* 13 (1982): 324–340.

Holmström, Bengt. “Pay for Performance and Beyond.” *American Economic Review* 107(7) (2017): 1753-1777.

Holmström, Bengt, and Paul Milgrom. "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, & Organization* 7 (1991): 24-52.

Hossain, Tanjim, and John A. List. "The behavioralist visits the factory: Increasing productivity using simple framing manipulations." *Management Science* 58.12 (2012): 2151-2167.

Ichniowki, Casey, and Kathryn Shaw. “Beyond Incentive Pay: Insiders’ Estimates of the Value of Complementary Human Resource Management Practices.” *Journal of the Economic Perspectives* 17.1 (2003):155-180.

Jovanovic, Boyan, and Yaw Nyarko. “Learning by Doing and the Choice of Technology.” *Econometrica* 64.6 (1996): 1299-1310.

Lazear, Edward P. “Performance Pay and Productivity.” *American Economic Review* 90.5 (2000): 1346-1361.

Lazear, Edward P. “Compensation and Incentives in the Workplace” *Journal of Economic Perspectives* 32.3 (2018): 195-214.

Shaw, Kathryn, and Edward P. Lazear. "Tenure and output." *Labour Economics* 15.4 (2008): 704-723.

Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent. "Performance pay and wage inequality." *Quarterly Journal of Economics* 124.1 (2009): 1-49.

Levitt, Steven D., and John List. “Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plan.” *Journal of Political Economy* 121.4 (2013): 643-681.

Levitt, Steven D., and Susanne Neckermann. “What field experiments have and have not taught us about managing workers.” *Oxford Review of Economic Policy* 30.4 (2015): 639-657.

List, John A., and Imran Rasul. "Field experiments in labor economics." *Handbook of Labor Economics*. 4. Elsevier (2011): 103-228.

Lourenço, Sofia M. “Monetary Incentives, Feedback, and Recognition – Complements or Substitutes? Evidence from a Field Experiment in a Retail Service Company.” *The Accounting Review* 91.1 (2016): 279-297.

Manso, Gustavo. “Motivating Innovation.” *Journal of Finance* 66 (2011): 1823-1860.

Prendergast, Canice. “The Provision of Incentives in Firms.” *Journal of Economic Literature* 37.1 (1999): 7-63.

Ravn, Morten, Stephanie Schmitt-Grohé, and Martin Uribe. “Deep habits.” *Review of Economic Studies* 73.1 (2006): 195-218.

Shearer, Bruce. “Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment.” *Review of Economic Studies* 71.2 (2004): 513-534.

Shi, Lan. “Incentive Effect of Piece-Rate Contracts: Evidence from Two Small Field Experiments.” *The B.E. Journal of Economic Analysis & Policy* 10.1 (2010): 1-32.

Weitzman, Martin L. “The "Ratchet principle" and performance incentives.” *The Bell Journal of Economics*, (1980): 302-308.

2.7 Appendix to Chapter 2

2.7.1 Additional Tables Experiment I

Table 2.4A – Balancing Table, Experiment 1

	(1) Descriptive Statistics	(2) *Norm. Bonus* District
Sales per Customer in October '15	12.8560 (1.5123)	0.568 (0.616)
Mean Sales per Customer '15	12.6136 (1.5138)	-0.599 (0.618)
Female District Manager (Y/N)	0.1633 (0.3734	-0.158 (0.210)
Store in City (Y/N)	0.8145 (0.2477)	-0.317 (0.443)
FTE	7.5433 (0.7056)	-0.118 (0.122)
Age of Store in Years	14.9901 (3.4515)	0.0362 (0.0254)
Store Space in m^2	746.5118 (44.0471)	-0.000664 (0.00223)
N of Observations	49	49
R^2		0.1049
F-Statistic		0.69 (p=0.6829)

Note: The table reports overall descriptive statistics (means and standard deviations) in column 1 and results from an ordinary least squares regression linear probability model in column 2. The dependent variable is a dummy variable equal to 1 if the manager is part of the treatment. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.5A – Robustness Check, Experiment 1

	(1)	(2)	(3)	(4)
		Sales per Customer		log (Sales per Customer)
	Single Difference	More T	Trimmed	FE
Treatment Effect	-0.0618	-0.0207	0.0092	-0.0010
Norm. Bonus	(0.4757)	(0.0475)	(0.0458)	(0.0027)
Time FE	Yes	Yes	Yes	Yes
District FE	No	Yes	Yes	Yes
District Manager FE	No	Yes	Yes	Yes
N of Observations	147	1225	611	637
N of Districts	49	49	48	49
Within R^2		0.9389	0.9562	0.9595
Overall R^2	0.1818	0.1289	0.1315	0.1197

Note: The table reports results from different estimations with sales per customer on the district level as the dependent variable in column 1-3 and the log value in column 4. Column 1 reports a single difference estimation with only the treatment months included and controlled for the mean average sales per customer of the last year. Column 2 increases the time period of the fixed effects regression by one year. Column 3 uses trimmed data in which every month the bottom and top 1% are dropped. Column 4 uses the log value of sales per customer instead of the absolute. All regressions control for possible refurbishments of a store. Robust standard errors are clustered on the district level and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.6A – Monthly Treatment Effects, Experiment 1

	(1)	(2)
	Sales per Customer	Sales per Customer
Treatment Effect 1st Month	-0.00171	-0.0205
	(0.0436)	(0.0444)
Treatment Effect 2nd Month	-0.0103	-0.0291
	(0.0903)	(0.0901)
Treatment Effect 3rd Month	0.0181	-0.0220
	(0.0384)	(0.0412)
Time FE	Yes	Yes
District FE	Yes	Yes
District Manager FE	No	Yes
N of Observations	637	637
N of Districts	49	49
Within R^2	0.9427	0.9478
Overall R^2	0.1043	0.1186

Note: The table reports results from fixed effects regressions with the sales per customer on the district level as dependent variable. The regressions account for time and district fixed effects and adds district manager fixed effects in column 2. The regressions compare pre-treatment observations (January 2015-October 2015) with the observation during the experiment (November 2015 – January 2016). *Treatment Effect* thus refers to the difference-in-difference estimator. All regressions control for possible refurbishments of a store. Robust standard errors are clustered on the district level and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.7A – Other Dependent Variables, Experiment 1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sales	Customers	Inventory Losses	Mystery Shopping	Ordering Up	Ordering Down	Sick Days
Treatment Effect	-0.0960	-0.0437	-0.140	0.0116	-0.0270	-0.0394	0.164
Norm. Bonus	(0.0656)	(0.0393)	(0.103)	(0.142)	(0.122)	(0.0859)	(0.192)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Manager FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N Observations	637	637	637	637	637	637	637
N of Distircts	49	49	49	49	49	49	49
Within R^2	0.8826	0.8103	0.7476	0.0803	0.2912	0.6167	0.2014
Overall R^2	0.2262	0.0362	0.5191	0.0001	0.2202	0.4095	0.0654

Note: The table reports results from fixed effects regressions with different standardized dependent variables on the district level. Column 1 and column 2 use sales and customers as the dependent variable, respectively. Column 3 has the known product waste (opposite to the unknown waste from, for example, theft) as the dependent variable. Column 4 uses a scoring done by mystery shoppers. Columns 5 and 6 use the percentage of upward (downward) corrections by the store managers to the ordering proposal as the dependent variable. The dependent variable in column 7 is the average number of sick days taken by employees in a store. The regression accounts for time, district, and district manager fixed effects. The regressions compare pre-treatment observations (January 2015-October 2015) with the observation during the experiment (November 2015 – January 2016. All regressions control for possible refurbishments of a store. Robust standard errors are clustered on the district level and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

2.7.2 Additional Tables Experiment II

Table 2.8A – Balancing Table, Experiment II

	(1) Descriptive Statistics	(1) *Simple Bonus*	(2) *Norm. Bonus*
Sales per Customer October ‘16	13.1854 (2.4626)	0.00658 (0.0155)	-0.0107 (0.0158)
Mean Sales per Customer ‘16	12.9382 (1.3389)	0.00687 (0.0267)	0.0136 (0.0272)
Female Store Manager (Y/N)	0.4366 (0.4968)	-0.0835 (0.0601)	0.0141 (0.0613)
Store in City (Y/N)	0.7852 (0.4114)	-0.00623 (0.0830)	-0.0501 (0.0847)
FTE	7.5583 (1.4900)	-0.000674 (0.0196)	0.0134 (0.0200)
Age of Store in Years	14.0385 (8.3681)	-0.00343 (0.00401)	0.000444 (0.00409)
Age of Manager in Years	38.9437 (9.6521)	-0.00502 (0.00380)	0.00450 (0.00387)
Tenure of Manager in Years	11.1409 (8.0818)	0.00390 (0.00465)	-0.00594 (0.00474)
Store Space in m^2	752.809 (106.804)	-0.000166 (0.000317)	-0.0000426 (0.000323)
Part of Exp I (Y/N)	0.5070 (0.5008)	-0.0111 (0.0568)	-0.0584 (0.0579)
N of Observations	284	284	284
R^2		0.0196	0.0168
F-Statistic		0.55 (p=0.8559)	0.47 (p=0.9114)

Note: The table reports overall descriptive statistics in column 1 (means and standard deviations) and results from an ordinary least squares regression linear probability model in column 2&3. The dependent variable is a dummy variable equal to 1 if the manager is part of the treatment Simple Bonus (column 2) or part of the treatment Norm. Bonus (column 3). 0 always refers to the control group. Note that for 10 observations we do not have date on job tenure. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.9A – Robustness Check, Experiment II

	(1)	(2)	(3)	(4)
		Sales per Customer		log (Sales per Customer)
	Single Difference	More T	Trimmed	FE
Treatment Effect	-0.0352	-0.0067	0.0077	0.0016
Norm. Bonus	(0.4043)	(0.0500)	(0.0469)	(0.0028)
Treatment Effect	0.2517	0.0372	0.0521	0.0029
Simple Bonus	(0.4428)	(0.0573)	(0.0552)	(0.0030)
Time FE	Yes	Yes	Yes	Yes
Store FE	No	Yes	Yes	Yes
District Manager FE	No	Yes	Yes	Yes
Store Manager FE	No	Yes	Yes	Yes
N of Observations	882	6729	3370	3473
N of Stores	294	294	290	294
Cluster	50	50	50	50
Within R^2		0.8081	0.8581	0.8670
Overall R^2	0.0719	0.0241	0.0365	0.0340

Note: The table reports results from different estimations with sales per customer on the store level as the dependent variable in column 1-3 and the log value in column 4. Column 1 reports a single difference estimation with only the treatment month included and controlled for the mean average sales per customer of the last year. Column 2 increases the time period of the fixed effects regression by one year. Column 3 uses trimmed data in which every month the bottom and top 1% are dropped. Column 4 uses the log value of sales per customer instead of the absolute value. All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. Robust standard errors are clustered on the district level and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.10A – Monthly Treatment Effects, Experiment II

	(1)	(2)
	Sales per Customer	Sales per Customer
Treatment Effect *Norm. Bonus* 1st Month	-0.0138 (0.0634)	0.0048 (0.0524)
Treatment Effect *Norm. Bonus* 2nd Month	-0.0184 (0.0492)	-0.0142 (0.0653)
Treatment Effect *Norm. Bonus* 3rd Month	-0.0427 (0.0396)	-0.0203 (0.0466)
Treatment Effect *Simple Bonus* 1st Month	0.100** (0.0485)	0.0978* (0.0565)
Treatment Effect *Simple Bonus* 2nd Month	0.00786 (0.0468)	0.0176 (0.0842)
Treatment Effect *Simple Bonus* 3rd Month	0.0166 (0.0764)	-0.0134 (0.0599)
Time FE	Yes	Yes
Store FE	Yes	Yes
District Manager FE	No	Yes
Store Manager FE	No	Yes
N of Observations	3822	3473
N of Stores	294	294
Cluster	50	50
Within R^2	0.8475	0.8478
Overall R^2	0. 0498	0. 0312

Note: The table reports results from a fixed effects regression with the sales per customer on the store level as the dependent variable. The regression accounts for time and district fixed effects and adds district manager fixed effects in column 2. The regressions compare pre-treatment observations (January 2016-October 2016) with the observation during the experiment (November 2016 – January 2017). *Treatment Effect* thus refers to the difference-in-difference estimator. All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.11A – Other Dependent Variables, Experiment II

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sales	Customers	Inventory Losses	Mystery Shopping	Ordering Up	Ordering Down	Sick Days
Treatment Effect	0.0280	0.0090	-0.0385	-0.0652	-0.0102	0.0097	0.0317
Norm. Bonus	(0.0435)	(0.0333)	(0.0616)	(0.0839)	(0.0765)	(0.0709)	(0.1320)
Treatment Effect	-0.0001	-0.0080	0.0615	-0.0078	0.0227	0.0054	-0.0244
Simple Bonus	(0.0407)	(0.0311)	(0.0676)	(0.1056)	(0.0808)	(0.0724)	(0.1053)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Manager FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Store Manager FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of Observations	3473	3473	3473	3472	3473	3473	3473
N of Stores	294	294	294	294	294	294	294
Cluster	50	50	50	50	50	50	50
Within R^2	0.6175	0.5537	0.4965	0.0407	0.1788	0.2719	0.0660
Overall R^2	0.0566	0.0040	0.2351	0.0098	0.0114	0.0749	0.0008

Note: The table reports results from fixed effects regressions with different standardized dependent variables on the store level. Column 1 and column 2 use sales and customers as the dependent variable, respectively. Column 3 has the known product waste (opposite to the unknown waste from, for example, theft) as the dependent variable. Column 4 uses a scoring done by mystery shoppers. Columns 5 and 6 use the percentage of upward (downward) corrections by the store managers to the ordering proposal as the dependent variable. The dependent variable in column 7 is the average number of sick days taken by employees in a store. The regression accounts for time, district, and district manager fixed effects. The regressions compare pre-treatment observations (January 2016 - October 2016) with the observation during the experiment (November 2016 – January 2017). All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. Robust standard errors are clustered on the district level and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.12A – Quantitative Questionnaire, Experiment II

	(1) Control	(2) Simple Bonus	(3) Norm. Bonus	(4) Difference (1)-(2)	(5) Difference (1)-(3)	(6) Difference (2)-(3)
The bonus formula was fair.		2.86 (1.74)	3.87 (1.68)			-1.001***
The bonus motivated me to raise my average receipt.		2.65 (1.65)	3.34 (1.7)			-0.691*
I tried to raise my average receipt in the previous months.	2.39 (1.45)	1.95 (0.82)	2.5 (1.43)	0.438*	-0.109	-0.547**
The bonus formula insures me against exogenous shocks.		3.28 (1.18)	3.84 (1.33)			-0.563**
The bonus depends on things I cannot influence.		2.88 (1.45)	2.34 (1.44)			0.542*
The size of the bonus was ok.		2.7 (1.5)	3.16 (1.37)			-0.460
I understood the bonus formula		2.07 (1.33)	3.55 (1.74)			-1.483***
The bonus formula was complicated.		4.56 (1.75)	2.79 (1.49)			1.769***
The average receipt can be influenced by store managers.	2.78 (1.36)	3.23 (1.25)	3.47 (1.29)	-0.450	-0.691**	-0.241
The average receipt can be influenced by district managers.	3.61 (1.48)	4.05 (1.38)	3.87 (1.34)	-0.438	-0.260	0.178
A store with an initially high average receipt can more easily influence the average receipt.	3.65 (1.62)	4.44 (1.26)	4.47 (1.29)	-0.790**	-0.822**	-0.032
A store with an initially low average receipt can more easily influence the average receipt.	2.65 (1.29)	3 (1.69)	3.05 (1.69)	-0.348	-0.400	-0.053
I know how to influence the average receipt.	2.39 (1.45)	2.12 (1.12)	2.32 (1.21)	0.275	0.076	-0.200
My district manager leaves me room to influence the average receipt.		3.23 (1.63)	3.47 (1.45)			-0.241
N Observations	53	43	38			

Note: The table reports means and standard deviations from the post-experimental questionnaire of experiment II. The questionnaire asked store managers to evaluate the statement on a scale from 1 (completely agree) to 6 (completely disagree). Column 4-6 report differences between treatment groups and statistical significance using a t-test. * p<0.1, ** p<0.05, *** p<0.01.

2.7.3 Prior Learning

Table 2.13A – Robustness Check, Prior Learning, Experience Proxy

	(1)	(2)	(3)	(4)	(5)
		Sales per Customer			log (Sales per Customer)
	Single Difference	More T	Trimmed Sales per Customer	Trimmed Experience Proxy	FE
Treatment Effect *Norm. Bonus*	0.728 (0.835)	0.302** (0.133)	0.238* (0.131)	0.299** (0.139)	0.0163** (0.00810)
Treatment Effect Norm. Bonus x *Experience Proxy*	-1.695 (1.919)	-0.590** (0.237)	-0.437* (0.232)	-0.590** (0.241)	-0.0277** (0.0132)
Treatment Effect *Simple Bonus*	1.845** (0.877)	0.331** (0.124)	0.231* (0.127)	0.332** (0.136)	0.0117 (0.00767)
Treatment Effect Simple Bonus x *Experience Proxy*	-3.494** (1.614)	-0.570** (0.227)	-0.343 (0.211)	-0.577** (0.244)	-0.0173 (0.0120)
Time FE	Yes	Yes	Yes	Yes	Yes
Store FE	No	Yes	Yes	Yes	Yes
District Manager FE	No	Yes	Yes	Yes	Yes
Store Manager FE	No	Yes	Yes	Yes	Yes
N of Observations	852	6526	3275	3315	3378
N of Stores	284	284	280	278	284
Cluster	50	50	50	50	50
Within R^2		0.8088	0.8584	0.8486	0.8669
Overall R^2	0.083	0.0274	0.0465	0.0388	0.0349

Note: The table reports results from different estimations with sales per customer on the store level as the dependent variable in column 1-4 and the log value in column 5. Column 1 reports a single difference estimation with only the treatment month included and controlled for the mean average sales per customer of the last year. Column 2 increases the time period of the fixed effects regression by one year. Column 3 uses trimmed data in which every month the bottom and top 1% of sales per customer are dropped. Column 4 uses trimmed data in which every month the bottom and top 1% of the experience proxy are dropped. Column 5 uses the log value of sales per customer instead of the absolute value. All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. Robust standard errors are clustered on the district level and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 2.14A – Heterogeneous Effects – Separate Experience Variables

	(1) Sales per Customer	(2) Sales per Customer
Treatment Effect Norm. Bonus	0.258* (0.129)	0.312** (0.139)
Treatment Effect Norm. Bonus x *Perc. Tenure Manager*	-0.252 (0.174)	-0.263 (0.179)
Treatment Effect Norm. Bonus x *Perc. Age Store*	-0.141 (0.172)	-0.133 (0.190)
Treatment Effect Norm. Bonus x *Perc. Age Manager*	-0.130 (0.168)	-0.210 (0.155)
Treatment Effect Simple Bonus	0.224* (0.130)	0.282* (0.143)
Treatment Effect Simple Bonus x *Perc. Tenure Manager*	-0.0930 (0.156)	-0.153 (0.152)
Treatment Effect Simple Bonus x *Perc. Age Store*	-0.148 (0.142)	-0.136 (0.174)
Treatment Effect Simple Bonus x *Perc. Age Manager*	-0.150 (0.153)	-0.205 (0.144)
Time FE x Percentile	Yes	Yes
Time FE	Yes	Yes
Store FE	Yes	Yes
District Manager FE	No	Yes
Store Manager FE	No	Yes
N of Observations	3692	3378
N of Stores	284	284
N of Cluster	50	50
Within R^2	0.8513	0.8531
Overall R^2	0.0485	0.0360

Note: The table reports results from a fixed effects regression with sales per customer on the store level as the dependent variable. The regression accounts for time and store fixed effects in column 1 and adds district manager and store manager fixed effects in column 2. The regressions compare pre-treatment observations (January 2016 - October 2016) with the observation during the experiment *TreatmentTime* (November 2016 – January 2017). All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. *Perc.* refers to the percentile of a store's age, manager's tenure, and manager's age of the respective manager/store. The regressions interact all time variables with store's age, manager's tenure, and manager's age. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

2.7.4 Instructions

2.7.4.1 Instructions Experiment I

Initial email to district managers in the bonus group (sent by regional manager)

Subject: Bonus "Average Receipt"

Dear Mr. XXX,

In the next three months, you can earn an additional bonus for increasing the average receipt in your district. For this, the monthly change of the average receipt in your district will be measured and you will be paid a bonus depending on this increase. The bonus will be calculated immediately after the end of a month and will be paid out to you at the beginning of the following month.

All district managers in the XXX region will receive an additional bonus in the time to come. However, due to administrative and evaluation-related reasons, the bonus will be paid out at two different points of time and will relate to two different performance measures. The two groups for this were randomly drawn according to a statistical method.

You are part of the first group and your three months bonus period starts on November 2nd, 2015. Therefore, we ask you to pay special attention to increasing the average receipt in the next months. Accordingly, the first bonus payment rewards an increase of the average receipt in the month November. Please consider the attached document for a more detailed explanation.

With kind regards

(Regional Manager)

Initial email to district managers in the control group (sent by regional manager)

Subject: Bonus "Average Receipt"

Dear Mr. XXX,

All district managers in the XXX region receive an additional bonus in the time to come. Due to administrative and evaluation-related reasons, the bonus period commences at two different points of time. The two groups for this were randomly drawn according to a statistical method. For fairness, the objective to increase the performance measure relates to two different performance measures. For the first group, the average receipt is relevant. The second group will learn its performance measure and its objective shortly before the beginning of the bonus period in the next year.

You are part of the second group and your bonus period starts next year. You will be informed about the exact period and the relevant performance measure at the beginning of the bonus period that is relevant to you.

With kind regards

(Regional Manager)

Attached Document to Email by Regional Manager

Initiative to Increase the Average Receipt

The average receipt is an essential driver of success for XXX. The aim of this initiative, which we are only conducting with half of the districts in your region due to administrative and evaluation-related reasons, is to increase the average receipt. The participating districts were randomly selected according to a statistical method.

Your district was selected. Therefore, we ask you to pay special attention to increasing the average receipt in the next months. Your success will be reported to you according to a performance measure on a monthly basis.

In the following month, you will receive this performance measure as a pay-out in Euros. The money will be credited to your employee card.

<u>Calculation of performance measure average receipt</u>

As of November 2nd, 2015, and until January 2016, you will receive monthly information regarding the increase of your average receipt. The performance measure depends on how the average receipt of your relevant stores[34] develops compared to the average receipt of the nation. I.e. the basis of calculation is:

Increase versus nation =

%-increase average receipt in the past month versus previous year (district)
– %- increase average receipt in the past month versus previous year (nation).

The increase is compared to a base value. The base value results from the comparison of the first three quarters of this year versus the first three quarters of the previous year.

Base value =

%-increase months 1-9 versus previous year (district)

– %-increase months 1-9 versus previous year (nation).

Hence, the base value stays the same for each month in which you receive information.

The performance measure is the difference between the increase versus nation and the base value. Therefore, it shows how your average receipt developed compared to the nation and the first months of the year.

Performance measure = (Increase versus nation – base value)*100

From this results a bonus pay-out of "€ performance measure"

[34] Relevant for the calculation are regular stores whose average receipt is not distorted by refurbishments.

2.7.4.2 Monthly Notifications Experiment I

Initiative to increase the average receipt

Monthly communication performance measure:

Dear Mr. XXX,

The first month of the project "Increase of Average Receipt" is now over. Listed below, you can find a summary of your average receipt figures.

Summary of your average receipt:

Increase versus nation

- Your average receipt in the last month was: XXX (X% increase versus previous year)
- The average receipt of the nation was XXX (X% increase versus previous year)

From this results an increase versus nation = %-increase receipt in the current month versus previous year (district)

– %-increase receipt in the current month versus previous year (nation)

= **XXX**

Constant base value

- Your average receipt from January to September this year was: XXX (X% increase versus previous year)
- The average receipt of the nation from January to September this year was XXX (X% increase versus previous year)

From this results a base value = %-increase months 1-9 versus previous year (district)

– %-increase months 1-9 versus previous year (nation)

= **XXX**

The resulting performance measure is: (Performance measure – constant base value) * 100 = XXX

Hence, we will credit € XXX to your employee card as soon as possible.

2.7.4.3 Instructions Experiment II

Initial letter to store managers in the bonus group (sent by regional manager)

Subject: Bonus "Average Receipt"

Dear XXX,

In the next three months, you can earn an additional bonus for increasing the average receipt in your store. For this, the monthly change of the average receipt in your store will be measured and you will receive a bonus depending on this increase. The bonus will be calculated immediately after the end of a month and will be paid out to you as part of the following payroll.

All store managers in the XXX region will receive an additional bonus in the time to come. However, due to administrative and evaluation-related reasons, the bonus will be paid out at two different points of time and might possibly relate to two different performance measures. The groups for this were randomly drawn according to a statistical method.

You are part of the first group and your three months bonus period starts on November 1st, 2016. Therefore, we ask you to pay special attention to increasing the average receipt in the next months. Accordingly, the first bonus payment rewards an increase of the average receipt in the month November. Please consider the attached document for a more detailed explanation.

With kind regards

(Regional Manager)

--

Initial letter to store managers in the control group (sent by regional manager)

Subject: Bonus "Average Receipt"

Dear Mr. XXX,

All store managers in the XXX region receive an additional bonus in the time to come. Due to administrative and evaluation-related reasons, the bonus commences at two different points of time. The groups for this were randomly drawn according to a statistical method. For fairness, the objective to increase the performance measure might possibly relate to two different performance measures. For the first group, the average receipt is relevant. The second group will learn about its performance measure shortly before the commencing bonus period in the next year.

You are part of the second group and your bonus period starts next year. You will be informed about the exact period and the relevant performance measure in the beginning of the bonus period that is relevant for you.

With kind regards

(Regional Manager)

Attached Document to Letter by Regional Manager (normalized bonus)

Initiative to Increase the Average Receipt

The average receipt is an essential driver of success for XXX. The aim of this initiative, which we are only conducting with two thirds of the store managers in your region due to administrative and evaluation-related reasons, is to increase the average receipt. The participating stores were randomly selected according to a statistical method.

Your store was selected. Therefore, we ask you to pay special attention to increasing the average receipt in the next months. Your success will be reported to you on a monthly basis according to a performance measure.

In the following month, you will receive this performance measure as a pay-out in Euros (capped upwards at € 375). The money will be credited to you as part of your payroll.

<u>Calculation of performance measure average receipt</u>

As of November 1st, 2016, and until January 31st, 2017, you will receive monthly information regarding the increase of your average receipt. The performance measure depends on how the average receipt of your store develops compared to the average receipt of the nation. I.e. the basis of calculation is:

Increase versus nation =

%-increase average receipt in the past month versus previous year (store)
– %-increase average receipt in the past month versus previous year (nation).

The increase is compared to a base value. The base value results from the comparison of the development in the first three quarters of this year versus the development in the first three quarters of the previous year.

Base value =

%-increase months 1-9 versus previous year (store)

– %-increase months 1-9 versus previous year (nation).

Hence, the base value stays the same for each month in which you receive information.

The performance measure is the difference between the increase versus nation and the base value. Therefore, it shows how your average receipt developed compared to the nation and the first months of the year.

Performance measure = (Increase versus nation – base value) * € 125

From this results a bonus pay-out of "€ performance measure"

<u>Fictitious example for normalized bonus</u>

<u>Increase versus nation</u>

Group	Avg.Receipt November 2016	Avg.Receipt November 2015	Increase in % versus prev. year	Increase in % versus nation
Store manager 1	13.78	13.51	2%	2% - 0.3% = **1.7%**
Nation	10.84	10.81	0.3%	

<u>Base value</u>

Group	Avg.Receipt cum. until Sept. 2016	Avg.Receipt cum. until Sept. 2015	Increase in % versus prev. year	Increase in % versus nation
Store manager 1	12.81	12.51	2.4%	2.4% - 2.7% = **- 0.3%**
Nation	10.28	10.01	2.7%	

Performance measure SM1 = (**1.7% - (- 0.3%)**) * 125 = 2 * 125 = 250

Performance measure in € = € 250

Attached Document to Letter by Regional Manager (simple bonus)

Initiative to Increase the Average Receipt

The average receipt is an essential driver of success for XXX. The aim of this initiative, which we are only conducting with two thirds of the store managers in your region due to administrative and evaluation-related reasons, is to increase the average receipt. The participating stores were randomly selected according to a statistical method.

Your store was selected. Therefore, we ask you to pay special attention to increasing the average receipt in the next months. Your success will be reported to you on a monthly basis according to a performance measure.

In the following month, you will receive this performance measure as a pay-out in Euros (capped upwards at € 375). The money will be credited to you as part of your payroll.

Calculation of performance measure average receipt

As of November 1st, 2016, and until January 31st, 2017, you will receive monthly information regarding the increase of your average receipt. The performance measure depends on how the average receipt of your store develops compared to the previous year. I.e. the basis of calculation is:

%-Increase average receipt in the past month versus previous year (store)

The performance measure is exactly this increase.

Performance measure = %-Increase versus previous year * € 125

From this results a bonus pay-out of "€ performance measure"

Fictitious example for simple bonus

Increase

Group	Avg.Receipt November 2016	Avg.Receipt November 2015	Increase in % versus prev. year
Store manager 1	13.78	13.51	**2%**

Performance measure SM1 = **2%** * 125 = 250

Performance measure in € = € 250

2.7.4.4 Monthly Notifications Experiment II

Normalized Bonus

Initiative to increase the average receipt

Monthly communication performance measure:

Dear Mr. XXX,

The first month of the project "Increase of Average Receipt" is now over. Listed below, you can find a summary of your average receipt figures.

Summary of your average receipt:

Increase versus nation

- Your average receipt in the last month was: XXX (X% increase versus previous year)
- The average receipt of the nation was XXX (X% increase versus previous year)

From this results an increase versus nation = %-increase receipt in the current month versus previous year (store)

– %-increase receipt in the current month versus previous year (nation)

= **XXX**

Constant base value

- Your average receipt from January to September this year was: XXX (X% increase versus previous year)
- The average receipt of the nation from January to September this year was XXX (X% increase versus previous year)

From this results a base value = %-increase months 1-9 versus previous year (store)

– %-increase months 1-9 versus previous year (nation)

= **XXX**

The resulting performance measure is: (Performance measure – constant base value) * 125 = XXX

Hence, we will credit € XXX as part of your next payroll as soon as possible.

Simple Bonus

<u>Initiative to increase the average receipt</u>

<u>Monthly communication performance measure:</u>

Dear Mr. XXX,

The first month of the project "Increase of Average Receipt" is now over. Listed below, you can find a summary of your average receipt figures.

<u>Summary of your average receipt:</u>

<u>Increase compared to previous year</u>

- Your average receipt in the last month was: XXX (X% increase versus previous year)

The resulting performance measure is: (Increase compared to previous year) * 125 = XXX

Hence, we will credit € XXX as part of your next payroll as soon as possible.

Chapter 3

3. Information Provision and Incentives – A Field Experiment on Facilitating and Influencing Managers' Decisions

3.1 Introduction

Standard principal-agent models assume that agents know the specific functional form linking effort to performance, but that their interests differ from the interests of their principal (employer). From this perspective, accounting information is used to assess and reward performance in order to align interests. However, if agents lack information about the underlying production function, this may limit their possibilities of yielding optimal work results even when interests are aligned. Hence, accounting scholars have stressed the dual role of information in organizations: information is used to *influence decisions* for instance through performance pay, and information serves to *facilitate decisions,* helping managers to make better decisions (Demski and Feltham 1976, Baiman 1982). The key purpose of this paper is to study the effect and interplay between these two roles of information using a firm-level field experiment.

Decision-facilitating information helps agents to make better decisions by providing ex-ante information to decrease uncertainty about specific actions and increase the agents' knowledge about the decision problem (Demski and Feltham 1976, Evans at el. 1994, Sprinkle 2003, Wall and Greiling 2011). Decision-influencing information, in contrast, is concerned with the evaluation of agents' behavior in order to affect their incentives through performance pay or supervisor monitoring (Demski and Feltham 1976, Baiman 1982). Research on principal-agent models has typically focused on this second role and studied the asymmetry of interests in organizations and the resulting control problems (see, e.g., Jensen and Meckling 1976, Holmström and Milgrom 1991, Merchant 1985, Sunder 1997, Indjejikian 1999, Prendergast 1999, Lazear 2000, Lazear 2018).

To understand the performance effects of both roles of managerial accounting information (facilitating and influencing) it seems essential to study the impact of both in the same environment as well as their interplay. As, for instance, put by Sprinkle (2003, p.288): "*It is important to study empirically how both roles of managerial accounting information affect the behavior of individuals who compose organizations*". To do so, we first analyze the interplay between decision-facilitating and decision-influencing information in a theoretical model. Then, we provide causal field evidence from a field experiment within the same organization on the effects of both roles of accounting information on performance and behavior.[1]

We implemented a firm-level field experiment with a 2x2 factorial design within a large German retail chain. The field experiment was conducted over a period of three months and 363 store managers of discount supermarkets within a particular geographical region of the retail chain were randomly assigned to three different treatment groups and a control group.

To facilitate their decisions, store managers in the "information" treatment group obtained information about the composition of profits and the underlying production function. The information consists of novel information about profit margins of individual products (which had not been known to store managers before), a short online training to brush-up knowledge on possible influences to increase profits, and a monthly electronic performance report concerning store profits.[2] Store managers in the "bonus" treatment group received monetary performance pay based on a simple profit metric. The third treatment group received both information and performance pay. The control group remains completely unaffected without any change or information about the other groups. The isolated and randomly assigned provision of either decision-facilitating information, decision-influencing with performance pay, and their combination allows to examine precise causal effects of the interventions (see e.g. Bandiera et al. 2011, Floyd and List 2016 for recent surveys on field experiments).[3]

We illustrate our key hypotheses by extending a standard multitasking moral hazard model (Holmström and Milgrom 1991, Baker 1992, Feltham and Xie 1994, Hemmer 1996) in which agents face uncertainty about the marginal returns to their efforts for different tasks. In the model, providing agents with information about those marginal returns helps managers to allocate efforts more efficiently across tasks. Furthermore, the benefits of this are larger when performance pay is in place such that our formal analysis suggests a complementarity between the two practices. We thus hypothesize that both, providing facilitating information and influencing decisions with performance pay alone, will increase profits.[4] We further

[1] While it has often been stressed that it is important to study interaction effects between different management practices (see, e.g., Milgrom and Roberts 1995, Holmström and Milgrom 1994, Ichniowski et al. 1997, Bonner and Sprinkle 2002, Hofmann and van Lent 2017), there is little clean causal evidence on such interdependencies (Lourenço 2016, Manthei et al. 2019, Sandvik et al. 2019).

[2] Prior to the intervention, store performance was mostly assessed by tracking single components of store profits such as sales and inventory losses. In a first step, a simple profit metric was introduced in all treatments as an aggregate accounting return measure. The online training and performance feedbacks helps to brush up the managers knowledge.

[3] For a discussion on endogeneity in managerial accounting research see Chenhall and Moers (2007) and Van Lent (2007).

[4] The experiment was pre-registered at the American Economic Association's registry for randomized controlled trials (AEARCTR-0002127). The experiments were approved by the workers council serving as an IRB substitute (as our institutions did not have an IRB at the time the experiment was carried out).

hypothesized that the combined implementation of information provision and performance pay will increase profits even more when the information helps managers to allocate their efforts more effectively.

Our formal model also highlights a close connection between the benefits of providing decision-facilitating information and the incentives to make use of this information: the provision of such information can only affect performance if the agent has an interest to act on it. While performance pay should strengthen such incentives, it is important to note that even without performance pay implicit incentives (or an employee's intrinsic identification with the firm's objectives) can generate the motivation to effectively make use of the provided information.

Our empirical results show the following: First, when introduced separately, both performance pay and the provision of decision-facilitating information have positive performance effects. Interestingly, the effect of information provision tends to outperform the effect of performance pay in nearly all specifications. This increase is driven by those managers who actually watched the training video. The combined intervention in which both practices are introduced together also increases performance. However, in contrast to our hypothesis, we find no evidence for a complementarity as the effect of the combined intervention never exceeds the sum of the separate effects of the two practices. Importantly, it achieves about the same effect on profits as the combined intervention. Moreover, when subtracting the costs of the bonus, the information provision treatment clearly outperforms the performance pay treatment, indicating that a lack of store managers' incentives to make use of provided information tends to be a less severe limiting factor than a lack of information. With a return of invest of approximately 5,500% the mere information intervention was also highly profitable.

In order to develop a more detailed understanding of the key result and the behavioral changes triggered by the interventions, we study data on sales and the number of sold products per category of product margins as well as questionnaire data. In particular, we developed questionnaires to elicit managers' task focus during the treatment period. We find that the provision of decision-influencing information substantially shifted the task focus towards *product placements*. Irrespective of whether the managers obtained performance pay or not, store managers put a substantially stronger emphasis on the placements of products when provided with information about the profit margins of individual products. Furthermore, we find that with the additional margin information store managers sell more products of higher margins. Hence, we find empirical evidence from multiple sources that the provision of decision facilitating information indeed changed managers' behavior and shifted their attention towards more profitable products.

Lastly, we study the effects on the single components of store profits, the persistence of the treatment effects, as well as the effect on managers' satisfaction. Here, we do not find evidence that the profit measure shifted the focus towards short-sighted actions (i.e. decreasing personnel costs). However, we find evidence in line with a crowding-out effect (Deci 1971, Gneezy et al. 2011, Huffman and Bognanno 2018) of financial incentives after the bonus is removed again, which is not present in the information treatment. Moreover, performance pay did not lead to greater satisfaction levels among store managers.

With these findings, we contribute to the literature on facilitating and influencing managers' decisions with information. There is empirical support for the benefits of both types of information but hardly any causal field evidence on their interplay or the effect of both information types in the same environment. The provision of decision-facilitating information can, for instance, increase learning and improve the quality of decisions (see, e.g., Ghosh 1997, Frederickson et al. 1999). Anderson and Kimball (2019) show that providing school teachers with information about students learning progress facilitates their diagnoses and possible focus to improve students' performance. In Casas-Arce et al. (2017a), the simple provision of customer lifetime value data to bank employees has a positive impact on the customer value and increases the employees' attention towards more profitable clients.[5] Manthei and Sliwka (forthcoming) show in a field experiment that providing supervisors with objective information of subordinates' performance raises profits in a retail bank. In a slightly different context, Farrell et al. (2008) show that contracting on a forward-looking measure facilitates effort allocation across multiple periods.[6]

The use of decision-influencing information in order to provide performance pay has also been the focus of a very broad empirical literature in accounting and economics. Most of the performance incentives and rewards that have been investigated have had positive effects on performance (see, e.g., Bailey et al. 1998, Banker et al. 2000, Sprinkle 2000, Lazear 2000, Shearer 2004, Presslee et al. 2013, Lourenço 2016, Friebel et al. 2017). Nevertheless, different circumstances such as task complexity, multitasking, different preferences, image concerns or

[5] The literature also shows some countervailing effects. For instance, too frequent (performance) information can reduce positive effects at least if the employees do not have the choice of receiving the information (Casas-Arce et al. 2017b, Holderness et al. 2019).

[6] The process of providing decision-facilitating information is also related to the use of trainings and knowledge transmission in organizations (see, e.g., Dearden et al. (2006) and Bassanini et al. (2007) for summaries of the training literature, and De Grip and Sauermann (2012) for a field experiment to estimate the effect of training on worker productivity). Work-related trainings can also be interpreted as filling a gap of knowledge (information) about a specific production function. Field experiments by Bloom et al. (2013) and Hanna et al. (2014) show, for instance, that managers are frequently not aware of the underlying production function and find substantial profit increases through the implementation of new practices.

exhausted learning curves can reduce the positive effect (e.g. Holmström and Milgrom 1991, Bonner et al. 2000, Frey and Jegen 2001, Bénabou and Tirole 2006, Sliwka 2007, Manthei et al. 2018).

Only a few studies directly look at the interplay between the two roles of information or their relative performance. In a laboratory experiment Sprinkle (2000), for instance, finds that feedback information to facilitate learning and the use of performance incentives are not independent and learning effects are greater when the provided information to facilitate learning is also part of the performance incentive. Using survey data, van Veen-Dirks (2010) shows that firms tend to attach more importance to the decision-facilitating rather than the decision influencing use of a broad set of accounting key figures. Grafton et al. (2010) argues that performance is correlated with the degree of commonality between the decision-facilitating and decision-influencing information.

We thus contribute to the literature by studying the relative importance and interplay of decision-influencing and decision-facilitating information in a firm-level field experiment. Having access to a unique dataset on financial KPIs as well as self-elicited survey data, we can investigate both monetary performance effects of the interventions but also changes in managers' behavior.

3.2 Hypothesis Development

As described in the above, the empirical literature on the provision of decision-facilitating information or using decision-influencing information to set incentives mostly shows that both practices have a positive effect on the performance of agents. We now adapt the classical framework of a multitasking principal agent model (see, e.g., Holmström and Milgrom 1991, Baker 1992, Feltham and Xie 1994, Hemmer 1996) to derive theoretical hypotheses about both the effect of decision-facilitating and decision-influencing information as well as their interplay within in the same theoretical framework.

3.2.1 A Simple Formal Model

We extend a standard Holmström and Milgrom (1991) type multitasking principal agent model to incorporate the role of decision-facilitating information by introducing uncertainty about the agent's marginal returns to their effort for different tasks. An agent is working on $j =$

$1, \ldots, k$ tasks and can exert a vector e of efforts e_j on task j at cost $\frac{1}{2}e_j^2$. Effort generates output for the principal where the marginal returns of effort are given by r_j for task j such that output is

$$\pi = \sum_{j=1}^{k} (e_j r_j + \eta_j)$$

where $\eta_j \sim N\left(0, \sigma_{\eta_j}^2\right)$ are independent noise terms. As in Bushman et al. (2000), marginal returns are ex-ante unknown and the r_j are independently drawn from a normal distribution with $r_j \sim N(m_j, \sigma_j^2)$.

We assume that the performance metric π is verifiable and can be used as decision-influencing information for performance pay. The agent receives a base pay α and may also obtain a bonus which is linear in π with slope β. We allow for the possibility that even without bonus pay the agent internalizes the principal's well-being to some extent (due to social preferences or implicit incentive from career concerns) such that her payoff is

$$h \cdot (\alpha + \beta \cdot \pi) + \theta \cdot \pi - \sum_{j=1}^{k} \frac{1}{2} e_j^2$$

where $h > 0$ measures the agent's marginal utility of money and $\theta \geq 0$ the degree to which he internalizes the effect of his actions on the principal's profits.

The agent has some prior information about the marginal productivities of the different tasks and observes a vector s of individual signals

$$s_j = r_j + \varepsilon_j$$

with $\varepsilon_j \sim N\left(0, \sigma_{\varepsilon_j}^2\right)$ for each task. The agent has a CARA utility function such that we can apply the standard result that her certainty equivalent is equal to

$$E[h \cdot (\alpha + \beta \cdot \pi) + \theta \cdot \pi | s] - \sum_{j=1}^{k} \frac{1}{2} e_j^2 - \frac{1}{2} r V[(\beta + \theta)\pi | s].$$

The agent maximizes[7]

[7] Note that due to the additive structure efforts do not affect risk premia.

$$\max_{e} h\alpha + E_A\left[(h\beta+\theta)\left(\sum_{j=1}^{k}\left(e_j r_j + \eta_j\right)\right)\middle| s\right] - \sum_{j=1}^{k}\frac{1}{2}e_j^2$$

and thus chooses

$$e_j = (h\beta+\theta)E_A\left[r_j \middle| s_j\right].$$

Hence, from the principal's perspective the ex-ante expected performance (without decision-facilitating information on the production function) is

$E\left[(\beta+\theta)\sum_{j=1}^{k} r_j\, E_A\left[r_j \middle| s_j\right]\right] = (h\beta+\theta)\sum_{j=1}^{k} E\left[r_j E_A\left[r_j \middle| s_j\right]\right].$

Using that $E_A\left[r_j \middle| s_j\right] = m_j + \sigma_j^2/(\sigma_j^2+\sigma_{\varepsilon_j}^2)\left(s_j - m_j\right)$ this becomes

$$(h\beta+\theta)\sum_{j=1}^{k} E\left[r_j\left(m_j + \frac{\sigma_j^2}{\sigma_j^2+\sigma_{\varepsilon_j}^2}\left(r_j+\varepsilon_j-m_j\right)\right)\right]$$

$= (h\beta+\theta)\sum_{j=1}^{k}\left(m_j^2 + \frac{\sigma_j^4}{\sigma_j^2+\sigma_{\varepsilon_j}^2}\right).$

We can now use this expression to study the impact of information provision and incentives on performance. To do so, suppose that the agent can receive additional decision-facilitating information that generates knowledge about the production function such that $\sigma_{\varepsilon_j}^2 = 0$, which implies that the agent learns the exact values of the marginal returns for each task. This directly implies:

Proposition: *The introduction of performance pay (i.e. choosing β>0), and the provision of decision-facilitating information (implementing $\sigma_{\varepsilon_j}^2 = 0$) both increase performance. The performance improvement from information provision*

$$(h\beta+\theta)\sum_{j=1}^{k}\sigma_j^2\left(1-\frac{\sigma_j^2}{\sigma_j^2+\sigma_{\varepsilon_j}^2}\right)$$

is larger when $\beta > 0$, such that performance pay and decision-facilitating information are complements.

Note that the model highlights a conceptually important point for the role of decision-facilitating information in affecting performance: decision facilitating information can only affect performance if there is some alignment of interest between principal and agent. If there is neither intrinsic alignment through employee identification or implicit incentives ($\theta = 0$) nor performance pay ($\beta = 0$), then decision facilitating information is useless as the agent has no incentives to act on it.

3.2.2 Hypotheses

The reviewed literature and our stylized model lead to the following hypotheses for our research setting. The first two hypotheses have been studied separately in the empirical literature in accounting and economics to a certain extent before. The third hypothesis is based on our formal model.

Hypothesis 1: The provision of information to facilitate decisions increases performance.

As illustrated by the formal model, the information about marginal productivities of the different tasks helps the agent to allocate efforts more efficiently across tasks as long as he has some incentives to make use of the information. This hypothesis is well in line with the empirical literature in accounting showing that decision-facilitating information through a variety of different channels tends to raise performance (frequent information (e.g. Frederickson et al. 1999), performance information (e.g. Holderness et al. 2019), novel information (e.g. Casas-Arce et al. 2017a)).

Hypothesis 2: Performance pay increases performance.

This hypothesis reflects just the standard incentive mechanism illustrated in moral hazard models: performance pay raises the agent's marginal returns of effort and thus increases these efforts. Most of the empirical literature in both accounting and economics on the causal effects of performance pay indeed supports the view that performance pay does have a positive influence on the agent's performance (e.g. Banker et al. 2000, Lazear 2000).[8]

[8] Limiting factors are, for instance, the lack of observability of important tasks (e.g. Holmström and Milgrom 1991), image concerns and motivation crowding-out (e.g. Bénabou and Tirole 2006) or exhausted learning curves (Manthei et al. 2018).

Hypothesis 3: The provision of information to facilitate-decisions and the use of performance pay are complements. That is, the impact of introducing performance pay is larger when decision-influencing information is provided and vice versa.

The key rationale for the hypothesis is illustrated in our formal model: when performance pay is in place, managers should have a stronger incentive to exert effort – that is the conflict of interest between principal and agent is reduced. When decision-facilitating information provides them with more precise information on marginal returns of different tasks, managers can more effectively allocate these efforts across tasks. In turn, the provision of decision-facilitating information should have a stronger effect on performance when the manager's and firm's interests are aligned to a stronger extent through performance pay.

But here it is important to note that the strength of the complementarity depends on the relative importance of explicit versus implicit incentives for the agent's behavior. If in our formal model the degree of prior interest alignment θ is large (for instance, because of substantial implicit incentives) relative to the marginal utility of money h, then the provision of decision-facilitating information raises profits also in the absence of performance pay. But the complementarity with the use of performance pay will then be weak.

3.3 The Empirical Setting

The company is a large, nationwide discount retailer operating supermarkets in Germany. The average sales area of 695 square meter and 6.6 fulltime equivalent employee (FTE). The average tenure of a store manager is 14.18 years.

In discount retailing, tasks and processes are typically highly standardized and store managers have only limited leeway in the procedures of the store. The central office determines, for example, the store layout, product choices and most of the placements of goods within stores. Store managers' duties are mainly operational tasks like taking care of the presentation of (fresh) products, refill of shelves, cleanliness of stores and efficient processes within the store (e.g. at the cashier desk). A computer system recommends order quantities based on an algorithm, but managers can overwrite the suggestions using their specific knowledge on local customer demand. They also have some leeway in temporary price reductions and special placements of goods within specific areas of the store. We classified the store managers' main tasks in Appendix 3.9A.

The next hierarchical level above the store manager is the level of the district managers. District managers are usually former store managers and manage about 6 stores per district. Store and district managers receive weekly and monthly electronic performance feedback from the accounting department of the company. On their computer, the store managers have access to their main KPI’s: sales, number of customers, average sales per customer, personnel hours, personnel costs, inventory losses overall, sales of fresh items, inventory losses of fresh items, availability of items and a mystery shopping score. Concerning these KPI the store managers see the absolute value of the week/month, the development with respect to the previous year, the development with respect to the planned KPI and the rank within the region. The same is visible for the accumulated values of the KPI over the year. Thus, the store and district managers receive regular and detailed electronic performance feedback which also allows district managers a close monitoring of store managers’ performance.

Prior to our study, the evaluation of performance of stores and store managers was mostly based on the components of store profits such as sales and inventory losses. As explained above, these figures are displayed in weekly and monthly reports to the store and district managers. One of the key conjectures arising from the discussions with the company was that the use of a broader profit metric should increase the scope for managers to raise performance (as, for instance, suggested in Bouwens and Van Lent 2007).[9] Store managers are used to analyze the components of profits due to their regular performance feedback. However, an issue at the outset was that the procurement prices for the goods sold are not publicly shared as low procurement prices constitute a central source for competitive advantage in (very price competitive) discount retailing. As store managers before our new intervention did not know the actual margins for different products precisely, their possibilities to raise profits were limited. Hence, we developed the idea to provide managers with information about profit margins which constitutes the key element of our decision-facilitating information treatments.

3.4 The Experiment

From April 2017 to June 2017, we randomly varied whether store managers received decision-facilitating information, performance pay or the combination of both among the 363

[9] Moreover, we decided to use the store’s planned value as a threshold for receiving a bonus and not solely the managers’ past performance). With this we reduce possible ratchet effects from using solely past performance (as for instance discussed in Bol and Lill 2015, Mahlendorf et al. 2015, Casas-Arce et al. 2017, see also Indjejikian et al. 2014).

stores in one region of the firm. For this we used a simplified profit metric. This key figure was computed as follows:

Store Profit = sales – costs of goods sold (cogs) – personnel costs – inventory losses.

Or even more simplified:

Store Profit = gross profit margin – personnel costs – inventory losses

The metric does not include costs that store managers cannot affect (such as e.g. store rents, costs of logistics, and overhead costs). Thus, we use one aggregated measure that entails all key elements of profits that can be affected by a store manager's actions to incentivize managers to use their full knowledge and set of possible actions.

In the performance pay treatments, store managers receive a bonus based on this key figure. In the information provision treatments, store managers got a brush-up concerning the performance metric and additionally received information on profit margins of all products.

3.4.1 Implementation

In total, we implemented four different treatment groups in a 2x2 factorial design, randomly varying at the same time whether store managers received performance pay (i.e. a bonus) and whether they obtained decision-facilitating information.

Table 3.1 - Treatments

		Decision-Facilitating	
		Information	No Information
Decision-Influencing	Bonus	N=92	N=88
	No Bonus	N=92	N=91

We used a stratified randomization (see, e.g., Athey and Imbens 2017) procedure depending on a prediction of the districts profits in the first treatment month. To construct the stratification groups, we use one year of past data through January 2017 and then predict profits for the district in April 2017 using a simple time-series model.[10] Within groups of four with similar predicted values, we randomly assigned the treatments. We randomized at the district

[10] Unfortunately, we had to randomize 3 months in advance as the data on profits come with a delay of one month and the central office required the group composition early to implement the required operational processes.

level (approx. 8 stores) to avoid possible spillover effects and confusion due to possible communication within districts. Table 3.10A in the Appendix shows summary statistics and balancing of treatment groups.[11]

Store managers of the treatment groups were notified about the respective treatment with a personalized letter sent to the address of their private home in the last week of March. The letter contained information about the treatment, which started on April 1st, 2017. Importantly, letters are in the corporate design of the company, signed by HR and the regional manager, and sent from the company's post office. The control group did not receive any notification. District managers were briefed in written form on how to react to questions concerning the experimental design.[12]

To complement the treatments, we also ran two large online surveys with store and district managers prior to and after the experiment. We sent personalized letters to their private home address in February 2017 as well as in the last week of June 2017. The letters contained an individual code managers had to enter online which allows us to match them to the other data. It was not possible to connect the surveys with the experiment.

During the whole time of the experiment neither the district nor the store managers knew that we as a university were involved in this project nor that the project was a designed experiment. The only time we communicated directly with the managers were the questionnaires. Here, we maintained the managers' anonymity as a research institute.

3.4.2 Treatment BONUS

Managers in this group received bonus payments based on the profit metric explained in the above. Bonuses were calculated as follows:

$$\text{Bonus (in €)}=[\text{Stores Profit} - (0.8 \cdot \text{Planned value of Stores Profit})] \cdot €0.05$$

Store managers, thus, receive €0.05 for every €1 profit they yield above a threshold of 80% of the planned value.[13] The planned value had been determined by the accounting department in the beginning of the year, based on a prediction algorithm. Bonuses are

[11] We detect some differences between treatment and control groups (although we were in charge of the randomization). The amount of differences should still be by chance, but controlling for these differences in a simple OLS regression leads to no notable differences in the treatment effects (see Appendix Table 3.10A). Moreover, differences are time constant and should not affect the fixed effects regressions.

[12] Exemplary letters to store and district managers are provided in the Appendix 3.10.2.

[13] Monetary profit shares are thus substantial as compared to, for instance, a usual CEO compensation of, according to Jensen and Murphy (1990), $3.25 for $1000 change in shareholder wealth. Moreover, the threshold is easy to reach. One month prior the experiment 90.27% of all stores exceeded the 80% of the planned profits threshold.

accumulated and cumulative bonuses are paid out after three months (capped at zero) together with the store managers' salary. Thus, in principle it is possible to receive a negative bonus for one month and lose part of the amount gained in previous bonus months. Note, that there were no individual performance bonuses available to store managers in this region before.

For each of the three months from April to June 2017, store managers in this treatment group also received a personalized letter sent to the address of their private home.[14] The letter reported the achieved profit and all its components of the previous month as well as the initially planned value. Moreover, managers received feedback on the bonus for the respective month.

3.4.3 Treatment INFORMATION

The provision of information to facilitate decisions consisted of an online training tool (a video explaining the profit metric and a quiz), information about the profit margin of individual products (which was not available to them before the intervention), and monthly electronic feedback on profits of the respective store. The online training tool was a 10-minute online video clip which explained the different profit components, how to influence them, and how they interact with each other.[15]

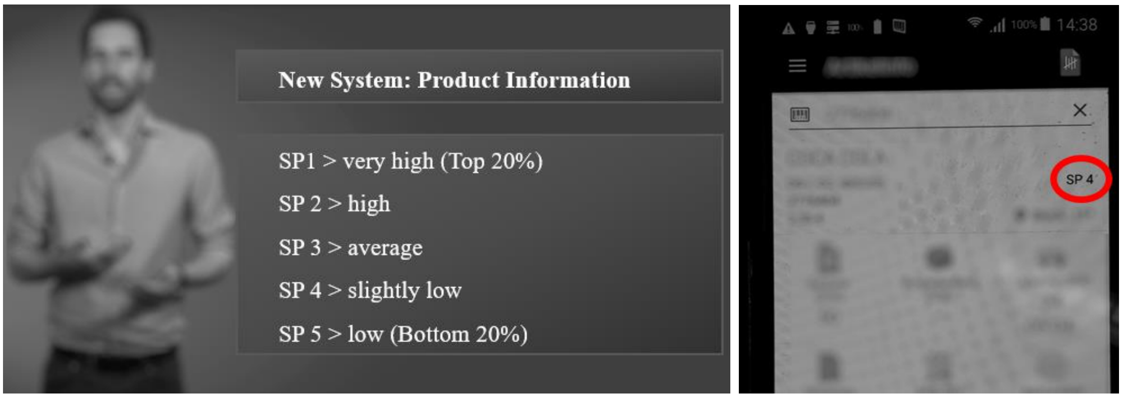
The video also explained the novel information managers obtain on profit margins in detail (see Figure 1 for a screenshot). As laid out in the above, the costs of goods sold for specific products are highly confidential in the competitive business of discount food retailing. Hence, the company had not disclosed specific margins prior to the experiment to store managers. In order to provide information about margins without giving precise information that could leak to competitors, we devised a system classifying all products according to their relative margin on a 5-point scale, where "1" meant that a product belonged to the 20% of products with the highest margins and "5" meant that it belonged to the quintile with the lowest margins. The intermediate steps were set accordingly. This margin rating was made accessible to store managers on their portable data terminals (PDT). PDTs are technical devices like smartphones with barcode scanners that are commonly used in retailing to immediately provide

[14] More precisely, due to a delay in calculating staff costs, the profit data was always delayed by one month. Hence, for instance, by the end of May we were able to send out the letter with the calculations for April. However, as explained in section 3 store managers receive their weekly and monthly electronic performance feedback from which they could directly infer how changes in their behavior led to changes in the financial KPI's. The letter is provided in the Appendix 3.10.2.

[15] As one of the authors was the trainer in the video clip and we scripted it, we had full control on the content and the transmission of the video. Store managers were not aware that the trainer was part of the research team. We carefully made sure that it remained a video to transfer and brush-up knowledge and not to motivate employees. A screenshots of the video is displayed in Figure 3.1. An excerpt of the video script is provided in the Appendix 3.10.3.

all product related information and allow for, for instance, quick ordering. Store managers thus had instant access to the information whenever scanning a product.

Figure 3.1 – Information on Profit Margins



Note: The left panel shows a screenshot from the video (pixelated and translated from German). The right panel shows the pixelated screen of the portable data terminal where the margin category was displayed (see circle).

The electronic performance report gave store managers information about the achieved profits for a given month as well as an overview on the components (sales, costs of goods sold, personnel costs, inventory losses) and planned values. It also showed the prior month's results along with the planned value for the next month. Moreover, the feedback reports also contained a reminder of the definition of the margin categories.

The key idea of this information intervention was thus to give store managers information about the store's production function and with this, facilitate a store manager's decision towards profit increases.

3.4.4 Treatment BONUS&INFORMATION

This treatment is a combination of individual monetary performance pay and information provision. It was conducted along the lines described above.

3.5 Results Experiment

3.5.1 Estimation

We estimate our main results on the full sample of managers originally assigned to the treatment (however, excluding managers who switched stores during the treatment time) using a difference-in-difference estimation including fixed effects for months and stores.

$$Y_{st} = \beta_0 + \beta_1 \cdot TreatmentBONUS_{st} + \beta_2 \cdot TreatmentINFORMATION_{st}$$
$$+\beta_3 \cdot TreatmentBONUS\&INFORMATION_{st} + \gamma \cdot X_{st} + a_s + \delta_t + \varepsilon_{st}$$

where Y_{st} is the profit in month t for store s. X_{st} includes time-variant controls, which here are the planned value of the store's profits as well as dummy variables indicating an ongoing or past refurbishment of the store. ε_{st} is an idiosyncratic error term clustered at the district level (the store belonged to at the beginning of the experiment). a_s are store fixed effects and δ_t are monthly time fixed effects. In some specifications, we also include district manager and store manager fixed effects. *TreatmentBONUS*, *TreatmentINFORMATION*, and *TreatmentBONUS&INFORMATION* are dummy variables equal to 1 for the respective treatment group during the experimental period and 0 otherwise.

We use the time periods from the beginning of the previous year to the end of the experiment (e.g. January 2016 through June 2017, 18 months) for the estimation of fixed effects. Table 3.11A in the Appendix provides the robustness checks with ordinary least squares regressions.

The key results are reported in Table 3.2. Column 1 displays outcomes of a fixed effects model of profits regressed on the treatment dummies, controlling planned values of the store profit and store refurbishments. Column 2 includes fixed effects for district managers and store managers. Columns 3&4 use the same specifications using logarithmized profits.

As laid out in the above, a key part of the INFORMATION treatments was the online training video in which also the margin categories where explained in detail. As we are able to track who took part in this online training, we also report regressions where we drop those store managers who did not watch the video. The respective results are displayed in Columns 5-8 of Table 3.2. We caution that training participation is affected by the treatment and thus endogenous. Hence, the estimates show the profit increases in the group of store managers that are sufficiently motivated to take part in the training.[16]

3.5.2 Main Results

On average, store managers with performance pay received a bonus payment of €969.99 at the end of the experiment (thus, on average €323.33 per month, which is approximately 10% of their monthly wage). In the first month of the experiment, only 3.31% of store managers (12 managers in total) did not reach the threshold of 80% of the planned profit and after the experiment only 1.93% did not receive a bonus at all.

The first key observation is that the provision of decision-facilitating information has a sizeable effect on profits irrespective of whether it is combined with a bonus or not. In fact, the *INFORMATION* treatment raised profits by about €1.000-€1.200 (about 2%) per month per store. If we drop observations of store managers who have not watched the online training video during the treatment period[17], the point estimates increase to about €1.300-€1500.[18] As the costs of the intervention were very small (costs of shooting the video and minor personnel costs of supplying the information), the intervention was highly beneficial for the firm with an approximate return on investment over the three months of the experiment of roughly 5500% just for the *INFORMATION* group (using the estimates from column 2).

While point estimates for the *BONUS* treatment are also positive, they tend to have a smaller magnitude. However, they are never statistically distinguishable from the effects of the information intervention (Wald test, p >0.1).

The combined information and performance pay intervention again yields economically significant point estimates of about €1400 per month and store. This does not exceed the estimates for the pure information intervention substantially. Studying the interplay between performance pay and the information intervention, we thus find no evidence for a complementarity between both practices. In all specifications the sum of the of the point

[16] Note that selection biases due to level differences in prior motivation are still accounted for in these specification as they all include store fixed effects and thus identify the effect from the within store variation in the treatment. Also recall that the video is only one part of the treatment as store managers in the information treatment also received performance reports. Hence, the information treatments may have affected store managers (and thus the dependent variable) not only through the learning video but also through the other components of the intervention. Therefore, we cannot estimate a local average treatment effect for compliers with the learning video alone as the exclusion restriction for an instrumental variable regression may be violated.

[17] For this specification we set store profit to missing during the treatment period in case a store manager did not watch the training video. Note that we also find a treatment difference in complying rates. In *INFORMATION* 80.43% of the store managers took part in the online training and 67.74% of *the BONUS&INFORMATION* group (MWU, p = 0.0495).

[18] These estimates have a causal interpretation (interpreted as the effect of the treatment on the store managers who are sufficiently motivated to watch the video), if we assume that the counterfactual time trends are uncorrelated with the motivation to take part in the online training,

estimates of *INFORMATION* and *BONUS* is larger than the *BONUS&INFORMATION* estimate. Using a Wald-test to test the sum of the isolated effects of *INFORMATION* and *BONUS* against the combined treatment effect of *BONUS&INFORMATION* does not yield a statistically significantly difference (Wald test, p >0.1) in any specification.

Thus, we find empirical support for hypothesis 1 and hypothesis 2. Providing information to facilitate decisions and using performance pay both do have a positive effect on agents' performance. But we do not find evidence for hypothesis 3 that the combined intervention increases profits more than the sum of both interventions implemented separately. Therefore, the two instruments are substitutes rather than complements.

Table 3.2 – Main Treatment Effects on Gross Profits

	Full sample				w/o managers who did not watch training video			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Profits	Profits	Ln (Profits)	Ln (Profits)	Profits	Profits	Ln (Profits)	Ln (Profits)
Treatment Effect BONUS	581.2	1050.2^{**}	0.0148	0.0276^{**}	581.2	1051.1^{**}	0.0149	0.0277^{**}
	(393.3)	(448.1)	(0.0115)	(0.0126)	(400.0)	(452.2)	(0.0116)	(0.0127)
Treatment Effect INFORMATION	998.6^{**}	1223.0^{**}	0.0173	0.0231	1324.9^{***}	1574.0^{***}	0.0281^{**}	0.0351^{**}
	(450.3)	(515.0)	(0.0143)	(0.0162)	(470.8)	(535.0)	(0.0137)	(0.0154)
Treatment Effect BONUS & INFORMATION	1295.3^{**}	1487.5^{**}	0.0257^{*}	0.0295^{*}	1292.7^{*}	1533.6^{**}	0.0241	0.0285
	(534.0)	(604.4)	(0.0137)	(0.0161)	(686.2)	(748.9)	(0.0167)	(0.0185)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Manager FE	No	Yes	No	Yes	No	Yes	No	Yes
Store Manager FE	No	Yes	No	Yes	No	Yes	No	Yes
Refurbishments	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Planned Profits	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wald test: BONUS=INFO	p=0.3526	p=0.7153	p=0.8525	p=0.7504	p=0.1146	p=0.2902	p=0.2881	p=0.5715
Wald test: BONUS+INFO= BONUS&INFO	p=0.6865	p=0.3235	p=0.7385	p=0.3258	p=0.4664	p=0.2376	p=0.3695	p=0.1379
N of Observations	6472	6296	6470	6296	6328	6154	6326	6154
N of Stores	363	363	363	363	362	362	362	362
Cluster	56	56	56	56	56	56	56	56
Within R^2	0.3492	0.3674	0.3246	0.3503	0.3514	0.3690	0.3305	0.3554
Overall R^2	0.8343	0.7479	0.7902	0.7102	0.8377	0.7564	0.8024	0.7210

Note: The table reports results from fixed effects regressions with the profits on the store level as the dependent variable. The regression accounts for time and store fixed effects (columns 1-8) and adds fixed effects for district and store managers in columns 2,4,6,8. Columns 5-8 drop observations for store managers who did not watch the training video during the treatment time. The fixed effects regressions compare pre-treatment observations (January 2016 - March 2017) with the observations during the experiment (April 2017 - June 2017). All regressions control for possible refurbishments of a store and the companies planned value. Observations are excluded once a store manager switched the store during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.001.

Table 3.3 replicates the same specifications when we use net profits that are adjusted by the respective costs of the treatments i.e. bonus payments are subtracted.[19] In this case, the dominance of the information intervention becomes all the more obvious. The provision of decision-facilitating information outperforms the combined intervention in each specification.

The results thus show that managers react to decision-facilitating information even in the absence of performance pay.[20] Moreover, the added value of using decision-influencing information through performance pay is rather small and we find no evidence that performance pay and information provision are complements. This seems at odds with the hypothesis developed using the formal model. However, as laid out in the above, if the agent's interests in our formal model are well aligned with the principal's objectives even in the absence of performance pay (in the model this corresponds to a high value of θ and a low value of h), the bulk of the effect of information provision is achieved even without performance pay and the value added from the complementarity becomes be small.[21] In other words, our results indicate that there are substantial implicit incentives for store managers to act on the provided information even without explicit incentives through performance pay.

[19] We do not consider indirect costs that result from more complex remuneration processes.

[20] An alternative interpretation for our results might be that the *INFORMATION* treatment affected the implicit incentives for store managers by communicating the importance of store profits relative to sales as key performance metric tracked by management without generating attention for the different profit margins. However, store profits were not a completely new KPI as store managers were already confronted with the components of store profits in their weekly/monthly reports. Moreover, the *BONUS* treatment should carry the same *signal* about the importance of store profits with an additional explicit incentive. Thus, incentives should be stronger in the BONUS treatment but in fact, point estimates are nearly always below those from the *INFORMATION* treatment. We explore this topic further in the next section when analyzing differences in managers' behavior.

[21] Recall that in our model the gain from information provision is $(h\beta + \theta)\sum_{j=1}^{k} \sigma_j^2 \left(1 - \frac{\sigma_j^2}{\sigma_j^2 + \sigma_{\varepsilon_j}^2}\right)$. When θ is large and h small this performance gain does not strongly depend on the size of the bonus β.

Table 3.3 – Main Treatment Effects on Net Profits

	Full sample				w/o managers who did not watch training video			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Profits	Profits	Ln (Profits)	Ln (Profits)	Profits	Profits	Ln (Profits)	Ln (Profits)
Treatment Effect BONUS	297.7 (392.7)	759.0* (445.1)	0.00914 (0.0155)	0.0234 (0.0170)	297.1 (399.1)	759.2* (449.0)	0.00923 (0.0157)	0.0236 (0.0170)
Treatment Effect INFORMATION	999.3** (450.5)	1224.1** (515.1)	0.0177 (0.0189)	0.0238 (0.0210)	1324.8*** (470.8)	1574.3*** (534.9)	0.0296* (0.0177)	0.0376* (0.0199)
Treatment Effect BONUS & INFORMATION	957.6* (512.5)	1146.9* (583.6)	0.0168 (0.0169)	0.0200 (0.0206)	941.1 (651.7)	1179.4 (714.0)	0.0143 (0.0204)	0.0187 (0.0231)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Manager FE	No	Yes	No	Yes	No	Yes	No	Yes
Store Manager FE	No	Yes	No	Yes	No	Yes	No	Yes
Refurbishments	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Planned Profits	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wald test: BONUS=INFO	p=0.1200	p=0.3251	p=0.6303	p=0.9834	p=0.0305	p=0.0996	p=0.2182	p=0.4204
Wald test: BONUS+INFO= BONUS&INFO	p=0.6216	p=0.2816	p=0.6822	p=0.3218	p=0.4024	p=0.1962	p=0.3489	p=0.1408
N of Observations	6470	6294	6468	6294	6326	6152	6324	6152
N of Stores	363	363	363	363	362	362	362	362
Cluster	56	56	56	56	56	56	56	56
Within R^2	0.3475	0.3656	0.3000	0.3240	0.3500	0.3675	0.3032	0.3263
Overall R^2	0.8350	0.7507	0.7729	0.7186	0.8381	0.6963	0.7954	0.7154

Note: The table reports results from fixed effects regressions with the net profits (profits minus bonus costs) on the store level as the dependent variable. The regression accounts for time and store fixed effects (columns 1-8) and adds fixed effects for district and store managers in columns 2,4,6,8. Columns 5-8 drop observations for store managers who did not watch the training video during the treatment time. The fixed effects regressions compare pre-treatment observations (January 2016 - March 2017) with the observations during the experiment (April 2017 - June 2017). All regressions control for possible refurbishments of a store and the companies planned value. Observations are excluded once a store manager switched the store during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.001.

3.6 Effects on Store Managers' Behavior

To analyze the effects and behavioral changes of our decision-facilitating and decision-influencing intervention in more detail, we make use of further detailed financial data as well as data from online questionnaires we have developed in order to assess the allocation of store managers efforts across tasks.

3.6.1 Task Focus

We invited store managers to participate in an online questionnaire close to the end of the experiment (204 store manager participated, participation rate 56.20%). This questionnaire contained closed questions on their task focus and job satisfaction. Moreover, we also included open questions asking store managers on what they did to increase profits and what they thought about the project (the information provision).

The task focus items asked store managers to evaluate the importance of specific tasks in their day-to-day work during the last 3 months (1 equals a low focus and 7 the highest focus). We categorized the tasks into seven task dimensions and display the mean of the different categories depending on the treatment in Figure 3.2 (the classification of tasks can be found in the Appendix Table 3.9A). Here, we detect no apparent treatments differences (see also the OLS regressions in Appendix Table 3.13A). Conceivably, when managers responded to these questions, they thought about their longer-term activities and routines and the between-task differences strongly dominate the within-task treatment differences.

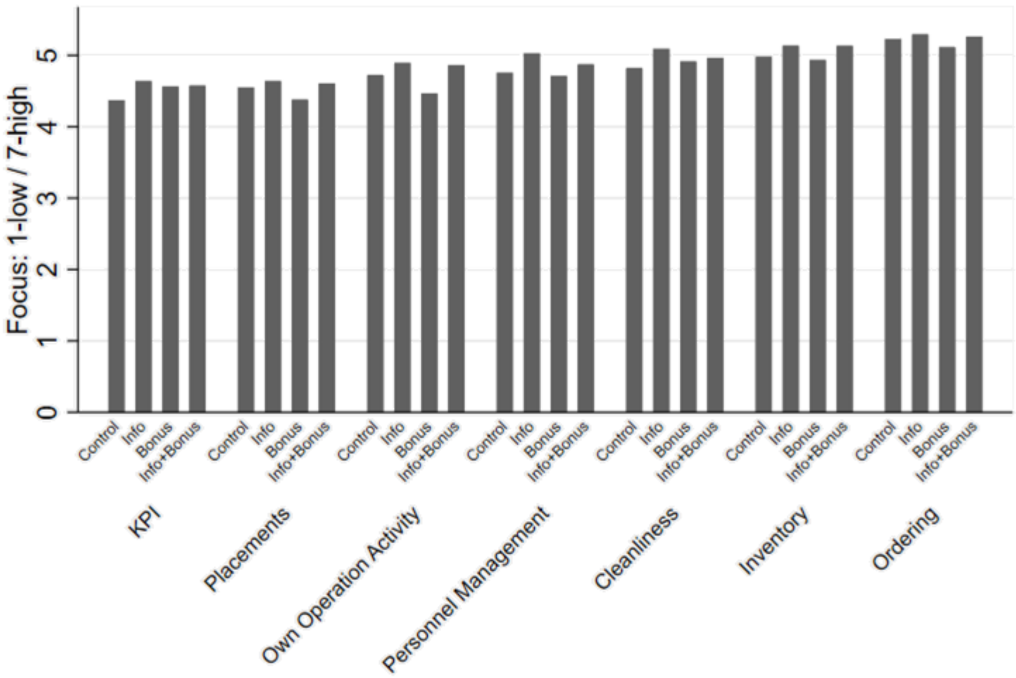
In a next step, we analyzed the answers to the open questions on what the managers did to increase profits in the months during the experiment. A research assistant classified the statements into the seven general task dimensions using the same categorization as before. The results are displayed in Figure 3.3. A first observation is that the relative order of tasks changes: While in their daily business the placements of goods does not play a dominant role (when ranking task dimension by order of importance as judged by the store managers it is the second to last of these seven dimensions – see Figure 3.3), it becomes the most important dimension when store managers are asked about activities implemented to raise profits. Frequently store managers stated that they made secondary placements of high margin products (products typically have specific locations in the store, but store managers can also display products on a second prominent spot – for instance on a specific desk close to the cash desk). Exemplary statements of store managers in the survey are for instance:

- *"I tried to prominently place articles with a high margin category (SP1 or SP2). Furthermore, I pushed sales of bakery products with secondary placements"*
- *"Paid attention to the margins of different articles and consequently made secondary placements"*
- *"Secondary placements in front of the cash desk. More focus on ordering of meat and bread"*
- *"Worked with the product margins, secondary placements of high margin products"*

Importantly, placements only stand out in the treatments with additional information on product margins. In fact, 38.30% of the survey respondents in the *INFORMATION* and 50% in the *INFORMATION&BONUS* group mentioned a placement activity while placements were only mention by 17.2% of respondents in the *BONUS* group (see also the regressions displayed in Table 3.4). The same picture arises when we only include statements that explicitly mention the placement of high margin products (See Figure 3.5A in the Appendix).

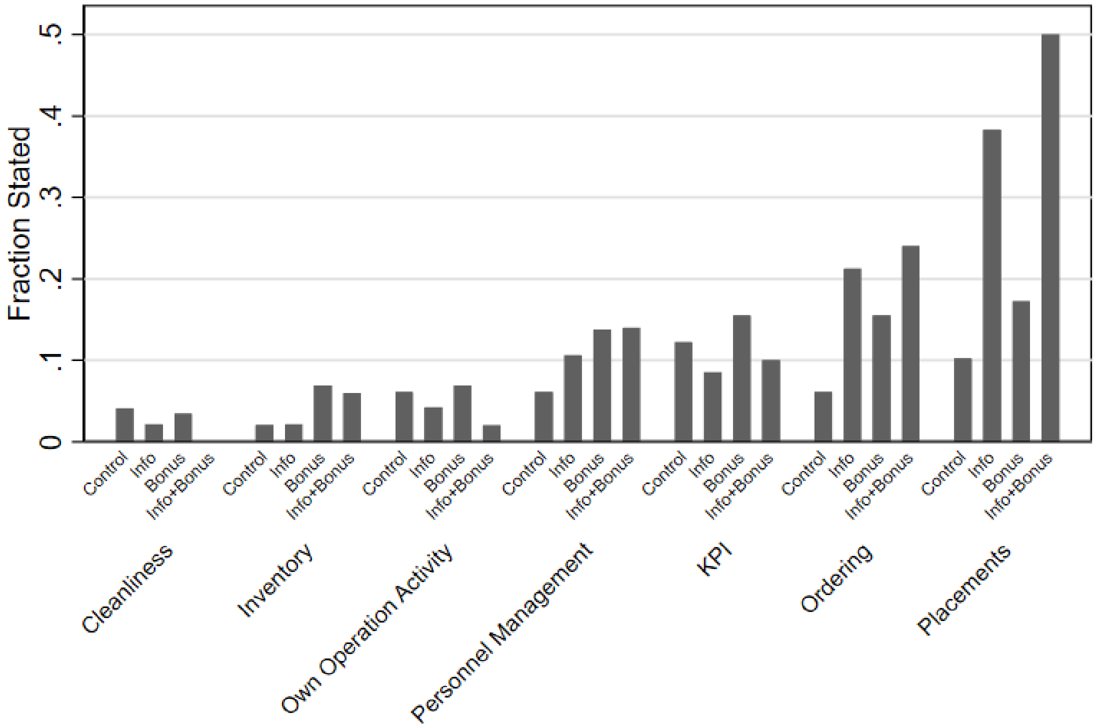
We observe a similar pattern for the managers' focus on product ordering. Hence, managers indeed reacted to the novel information on profit margins and they did so in particular with respect to ordering and the placement of high margin products. The control group, which remains completely unaffected in this study, state the lowest amount of activities to increase profits for all categories. But, interestingly, also managers in this group stated some activities to increase profits indicating that thinking in terms of profits was not completely new to the managers and also part of their daily routine.

Figure 3.2 – General Task Focus (Closed Questions)



Note: The figure displays the average rating of focus on specific tasks (1=low focus,6=high focus) obtained from an online questionnaire. Tasks were clustered into 7 dimensions. N=204.

Figure 3.3 – Task Focus to Increase Profits (Open Questions)



Note: The figure displays the fraction of stated tasks dimensions to increase profits obtained from open questions of an ex-post questionnaire. N=204.

Finally, it is also interesting to consider what managers did in the bonus group (relative to the control group). But note that here we do not detect any sizeable shifts in the task focus which may indicate that store managers in the bonus treatment just increased the intensity of their efforts without strongly shifting attention towards specific task dimensions.[22]

[22] Note that in the formal model, under symmetric prior information on tasks, the bonus should increase performance in all task dimensions similarly.

Table 3.4 - Self-Stated Actions to Increase Profits (Open Questions)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A	Ordering	Placements	Cleanliness	KPI	Inventory	Personnel Management	Own Effort
Treatment Effect BONUS	0.0939 (0.0718)	0.0704 (0.0827)	-0.00633 (0.0301)	0.0327 (0.0629)	0.0486 (0.0400)	0.0767 (0.0617)	0.00774 (0.0421)
Treatment Effect INFORMATION	0.152^{**} (0.0756)	0.281^{***} (0.0871)	-0.0195 (0.0317)	-0.0373 (0.0662)	0.000868 (0.0421)	0.0452 (0.0649)	-0.0187 (0.0443)
Treatment Effect INFORMATION &BONUS	0.179^{**} (0.0744)	0.398^{***} (0.0857)	-0.0408 (0.0312)	-0.0224 (0.0652)	0.0396 (0.0414)	0.0788 (0.0639)	-0.0412 (0.0437)
Controls	No	No	No	No	No	No	No
N Observations	204	204	204	204	204	204	204
R^2	0.033	0.124	0.010	0.007	0.012	0.010	0.008
Panel B	Ordering	Placements	Cleanliness	KPI	Inventory	Personnel Management	Own Effort
Treatment Effect BONUS	0.0682 (0.0844)	0.0287 (0.0948)	-0.0553^{*} (0.0285)	0.0450 (0.0716)	0.0295 (0.0459)	0.122 (0.0747)	-0.0164 (0.0489)
Treatment Effect INFORMATION	0.162^{*} (0.0834)	0.274^{***} (0.0937)	-0.0233 (0.0282)	-0.0713 (0.0707)	-0.00866 (0.0454)	0.0550 (0.0739)	-0.0245 (0.0483)
Treatment Effect INFORMATION &BONUS	0.160^{*} (0.0819)	0.373^{***} (0.0920)	-0.0440 (0.0277)	-0.0700 (0.0694)	0.0259 (0.0446)	0.0772 (0.0725)	-0.0443 (0.0474)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N Observations	181	181	181	181	181	181	204
R^2	0.060	0.165	0.040	0.040	0.040	0.032	0.028

Note: The table reports results from OLS regressions with the focus of different classified tasks from an online questionnaire as the dependent variable. The dependent variable is equals 1 if a mentioned task falls into the respective category and 0 otherwise. Panel B controls include the size of the store, amount of full time equivalent employees (FTE), age of the store manager, and the annual subjective performance evaluation. Observations are excluded once a store manager switched the store during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses.* p<0.1, ** p<0.05, *** p<0.001.

3.6.2 The Treatment Effects on Sales by Margin Category

As outlined above, store managers have set a greater focus on product placements in general and especially on placements of high margin products when they had access to the product margin information.

We additionally have access to data on sales and number of products sold for each of the five margin categories explained in section 3.4.3. In order to explore the treatment effects on sales by category, we regress these variables in the different margin categories separately on the treatment dummies. The results are displayed in Table 3.5. In line with the idea that profit increases are driven through the novel information on product margins, the *INFORMATION* treatment tends to lead to higher sales in the top but not the bottom margin categories (Panel A).[23] A similar pattern can be observed in the point estimates of the *BONUS&INFORMATION* treatment, but here standard errors are substantially larger such that the statistical power is too low to discern the effects.

We see a similar picture for the number of actual products sold within the margin categories (Panel B) providing further support for the claim that store managers actually focused on the revealed individual product margins when presenting and selling items within the store.

[23] Note, that it might be perfectly rational not to focus on the very best product margin category as these products are not necessarily the most popular products for customers.

Table 3.5 – Treatment Effects Depending on Product Margin Categories

Panel A –Sales	(1)	(2)	(3)	(4)	(5)
	1st Cat	2nd Cat	3rd Cat	4th Cat	5th Cat
Treatment Effect BONUS	132.9 (462.1)	101.6 (283.0)	291.7 (414.8)	-469.9 (918.1)	62.61 (307.3)
Treatment Effect INFORMATION	526.6 (409.4)	568.3** (251.1)	976.6** (382.5)	4.1 (919.7)	353.7 (260.3)
Treatment Effect BONUS & INFORMATION	586.0 (571.4)	576.8 (443.1)	980.2* (568.9)	444.8 (1185.4)	390.9 (323.4)
Fixed Effects (Store, Time, Store and District Manager)	Yes	Yes	Yes	Yes	Yes
Refurbishments, Planned Values	Yes	Yes	Yes	Yes	Yes
N Observations	6297	6297	6296	6297	6296
N Store	363	363	363	363	363
Cluster	56	56	56	56	56
Within R^2	0.5969	0.5615	0.5773	0.7913	0.8646
Overall R^2	0.7812	0.8578	0.8012	0.7480	0.7175
Panel B – Amount of Products	(1)	(2)	(3)	(4)	(5)
	1st Cat	2nd Cat	3rd Cat	4th Cat	5th Cat
Treatment Effect BONUS	-62.30 (537.8)	199.0 (157.9)	316.3 (324.5)	-329.4 (614.3)	10.87 (145.7)
Treatment Effect INFORMATION	423.5 (498.1)	383.4** (172.5)	885.8*** (328.2)	29.52 (605.6)	132.7 (127.7)
Treatment Effect BONUS & INFORMATION	957.0* (565.3)	614.9* (325.4)	1234.5** (597.9)	814.1 (636.5)	186.4 (163.3)
Fixed Effects (Store, Time, Store and District Manager)	Yes	Yes	Yes	Yes	Yes
Refurbishments, Planned Values	Yes	Yes	Yes	Yes	Yes
N Observations	6264	6264	6263	6264	6263
N Store	361	361	361	361	361
Cluster	56	56	56	56	56
Within R^2	0.4938	0.4469	0.3975	0.6116	0.8309
Overall R^2	0.0329	0.0525	0.0313	0.0667	0.2156

Note: The table reports results from fixed effects regressions with sales on the store level as the dependent variable. The different columns represent the different margin categories (e.g. 1st category = sales of the 20% of the products with the highest margin, 5th category = sales of the 20% of the products with the lowest margin). The regression accounts for time, store, store manager and district manager fixed effects. The fixed effects regressions compare pre-treatment observations (January 2016 - March 2017) with the observations during the experiment (April 2017 - June 2017). All regressions control for possible refurbishments of a store and the companies planned value for all profit components. Observations are excluded once a store manager switched the store during the treatment period. Observations are further excluded for store managers who did not watch the training video during the treatment time. *Treatment Effect* thus refers to the difference-in-difference estimator. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.001.

3.7 Additional Results

This section presents further results, for instance exploring the composition of the profit effects and studying the effects after the end of the experiment (after the bonus has been withdrawn again).

3.7.1 Other Financial KPIs

In a next step, we decompose the overall effect on our considered profit metric into

Store Profit = gross profit margin – personnel costs – inventory losses.

As this metric is linear and additive in the components, we can run regressions on each component separately and the sum of the effects on the components will roughly correspond to the overall effect of the treatment on the profit metric.[24] Hence, we analyze the treatment effects on the different components of profit running separate regressions for each of them. The results are displayed in Table 3.6.

We do not see highly significant treatment differences as standard errors are large which to some extent maybe due to the fact that treatment assignment was not stratified with respect to these key figures. But the point estimates indicate that the bulk of the effect in the information treatments is driven by an increase in the gross profit margin which again supports the conjecture that the provision of information helped store managers to allocate their attention towards more profitable sales.

Moreover, we find no evidence that the use of profit-based performance metrics lead to short-term thinking. While personnel costs and inventory losses are mainly lower in the treatment groups as compared to the control group, these savings apparently did not lead to losses in sales. The interventions thus appear to have supported sustainable top-line growth rather than short-term gains by cutting personnel costs or inventory losses.

[24] Note that the summation of the single effects will only roughly correspond to the overall main effect due to the introduction of fixed effects on various levels.

Table 3.6 – Treatment Effects on Profit Components

	(1)	(2)	(3)	(4)	(5)	(6)
	Gross Profit Margin	Personnel Costs	Inventory Losses	Gross Profit Margin	Personnel Costs	Inventory Losses
Treatment Effect BONUS	-58.52 (459.8)	-344.7 (280.2)	-73.31 (155.0)	487.4 (447.9)	-266.1 (220.6)	-96.91 (163.8)
Treatment Effect INFORMATION	609.7 (484.3)	-185.3 (284.7)	-157.1 (172.5)	954.9* (482.2)	-136.5 (224.9)	-128.3 (186.1)
Treatment Effect BONUS&INFORMATION	1059.6 (671.6)	101.8 (343.6)	-182.1 (131.4)	1350.4* (699.1)	207.8 (322.4)	-172.1 (164.9)
Fixed Effects (Store and Time)	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects (Store and District Manager)	No	No	No	Yes	Yes	Yes
Refurbishments	Yes	Yes	Yes	Yes	Yes	Yes
Planned Values	Yes	Yes	Yes	Yes	Yes	Yes
N Observations	6475	6475	6475	6298	6298	6298
N Store	363	363	363	363	363	363
Cluster	56	56	56	56	56	56
Within R^2	0.5861	0.4591	0.1146	0.5900	0.4795	0.1265
Overall R^2	0.9255	0.7593	0.1344	0.8750	0.6892	0.0529

Note: The table reports results from fixed effects regressions with different dependent variables on the store level as the dependent variable. All regressions account for time, store, store manager and district manager fixed effects. Furthermore, all regressions control for the planned product margin, personnel costs and inventory losses. The fixed effects regressions compare pre-treatment observations (January 2016 - March 2017) with the observations during the experiment (April 2017 - June 2017). All regressions control for possible refurbishments of a store and the companies planned value for the respective component. Observations are excluded once a store manager switched the store during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.001.

3.7.2 Persistence

In a next step, we explore how the treatment effects develop over time. Here we interact dummies for the months during the treatment, as well as the two months after the treatment, separately with our treatment groups using our main specification. The regression estimates are shown in Table 3.12A in the Appendix with specification along the lines of Table 3.2. Figure 3.4 displays the regression results of column 2 (Table 3.12A) with the full set of fixed effects. *BONUS* and *INFORMATION* show a substantial increase in the first month of the experiment. However, the effect gets weaker over the three month period.[25] While in the last month of the experiment the treatment effect in *INFORMATION* is not statistically significantly different

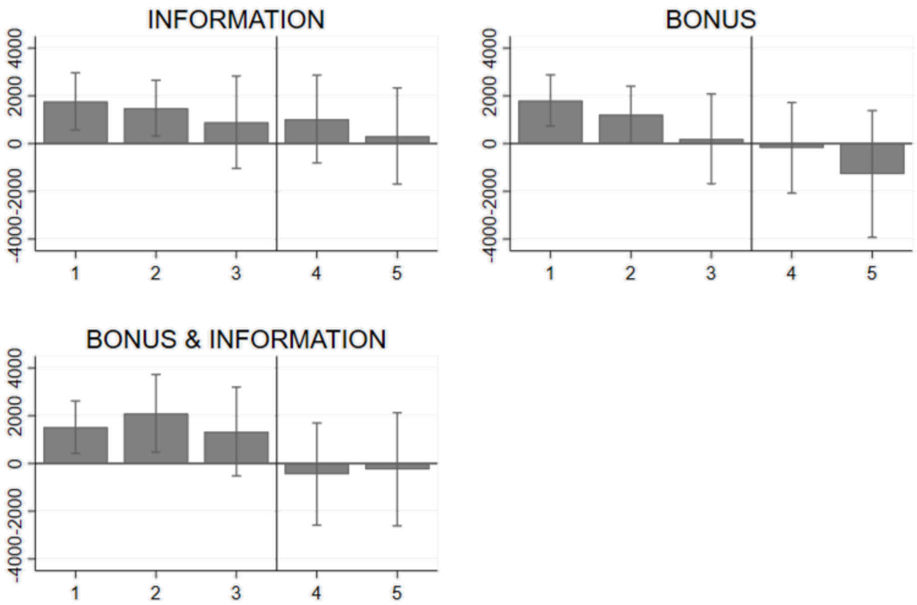
[25] Note, that this development is in general not specific to our setting. The treatment effect in De Grip and Sauermann (2012), for instance, also decreases over time. They explain this pattern by positive spillovers of trained/informed workers to untrained/uninformed workers in the control group. However, one could in principal argue that the intervention only increased attention on the key figure in the first month. Focusing attention has, for example, recently been studied by Gosnell et al. (forthcoming). In this study the authors provide a group of airline pilots with information about their flight behavior and find an increase in their flight performance.

from the first treatment month (Wald test, p = 0.3252, Table 3.12A column 2), the *BONUS* treatment shows a drop in the effect size between the first and the third treatment month (Wald test, p = 0.0926, Table 3.12A column 2). The decreasing treatment effect is in line with an attention-directing effect that vanishes over the time period (see, e.g., Simon 1954, Mock 1971, Hirshleifer and Teoh 2003). For the information treatment, this effect might have been expected as managers' awareness of the profit metric and its implications may naturally fade the longer the time passed since the initial information intervention (in our treatment the video). However, for the performance pay treatment, the effect seems at odds with standard incentive theory: the marginal incentives to raise profits in month one of our treatment were the same as in month 2 and 3. However, the treatment effects were smaller in later months. Apparently, one part of the effect of performance pay was to generate attention for the underlying performance metric and this attention was reduced over the time of the treatment.

Interestingly, treatment effects drop after the bonus is removed in July (4th month) and point estimates even become negative. The regression output in Table 3.7 shows the treatment effects during the treatment period, but also separately shows the effects after the intervention. In the two months after the intervention, the effects of both bonus groups (*BONUS* and *BONUS&INFORMATION*) drop below the control group with significant differences to the treatment month' as well as to the *INFORMATION* group after the experiment. Although speculative, this might be evidence for motivational crowding-out after removing a financial bonus (Deci 1971, Gneezy et al. 2011, Huffman and Bognanno 2018).[26]

[26] An alternative explanation is that store managers intentionally shifted profits forward into the treatment time frame. However, note that in supermarkets the strategic shifting of customers to different month is difficult if not impossible. Indirect shifting of profits by reducing costs in the treatment time at the expense of sales seems possible, but, as shown in section 3.7.1. we do not find evidence for such a behavior.

Figure 3.4 – Development of Treatment Effects on Profits



Note: The figure displays the average treatment effects on store profit resulting from our fixed effects specification in Appendix Table 3.12A (which is the specification of Table 3.2 column 2 and interacted with month during and after the treatment). 1 on the x-axis refers to the first treatment month and the vertical line indicated the end of the treatment. 95% confidence bands are displayed.

Table 3.7 – Treatment Effects During and After Intervention

	(1) FE	(2) FE
Treatment Effect BONUS	581.2 (405.63)	1045.02** (467.68)
Treatment Effect INFORMATION	1039.05** (460.0)	1350.68** (531.68)
Treatment Effect BONUS & INFORMATION	1359.45** (531.31)	1552.95** (610.59)
Post-Treatment Effect BONUS	-1217.13 (802.95)	-767.69 (867.31)
Post-Treatment Effect INFORMATION	292.2 (672.10)	654.10 (720.42)
Post-Treatment Effect BONUS & INFORMATION	-932.77 (875.94)	-490.95 (906.89)
Time FE	Yes	Yes
Store FE	Yes	Yes
District Manager FE	No	Yes
Store Manager FE	No	Yes
Refurbishments	Yes	Yes
Planned Profits	Yes	Yes
Wald test: BONUS=Post-BONUS	p=0.0121	p=0.0124
Wald test: INFORMATION=Post-INFO	p=0.1440	p=0.1757
Wald test: BONUS&INFO=Post-BONUS&INFO	p=0. 0024	p=0.0030
Wald test: Post-BONUS=Post-INFO	p=0.0605	p=0.0968
N of Observations	7196	6296
N of Stores	363	363
Cluster	56	56
Within R^2	0.3344	0.3674
Overall R^2	0.8283	0.7479

Note: The table reports results from fixed effects regressions with the profits on the store level as the dependent variable. The regression accounts for time and store fixed effects and adds fixed effects for district and store managers in column 2. The fixed effects regressions compare pre-treatment observations (January 2016-March 2017) with the observations during the experiment (April 2017 - June 2017). Moreover, "Post-Treatment" refers to the observations after the experiment (July 2017 - August 2017). All regressions control for possible refurbishments of a store and the companies planned value. Observations are excluded once a store manager switched the store during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.001.

3.7.3 Project Assessment by Store Managers

In our online questionnaire, we asked further questions, for example regarding the store managers' satisfaction. Table 3.8 displays OLS regressions using the survey measures of satisfaction as independent variables. Satisfaction is divided into satisfaction with the job (columns 1&4), satisfaction with compensation (columns 2&5), and satisfaction with the workload (columns 3&6). We find that satisfaction among store managers does not significantly vary across treatment groups. Job satisfaction in the bonus group is weakly significantly larger than in the control group, but there is no sizeable difference to the two other treatment groups.

Table 3.8 – Store Managers Satisfaction

	(1) Job	(2) Compensation	(3) Workload	(4) Job	(5) Compensation	(6) Workload
Treatment Effect BONUS	0.325 (0.243)	-0.00598 (0.222)	0.118 (0.243)	0.418* (0.240)	0.169 (0.223)	0.0867 (0.283)
Treatment Effect INFORMATION	0.349 (0.285)	-0.0337 (0.270)	0.349 (0.240)	0.253 (0.245)	-0.0462 (0.250)	0.244 (0.266)
Treatment Effect BONUS & INFORMATION	0.313 (0.250)	-0.116 (0.278)	-0.120 (0.268)	0.317 (0.226)	-0.017 (0.206)	-0.187 (0.251)
Controls	No	No	No	Yes	Yes	Yes
N Observations	203	203	203	179	179	179
R^2	0.019	0.001	0.017	0.125	0.135	0.114

Note: The table reports results from OLS regressions with different satisfaction outcomes obtained from an online questionnaire as the dependent variable. Column 4-6 control include the size of the store, amount of full time equivalent employees (FTE), age of the store manager, and the annual subjective performance evaluation. Observations are excluded once a store manager switched the store during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses.* p<0.1, ** p<0.05, *** p<0.001.

3.8 Conclusion

We report a firm-level field experiment studying the hypotheses that: (i) the provision of decision-facilitating information raises profits, (ii) using decision-influencing information by introducing of performance pay raises profits and (iii) both practices are complements. In order to do so, we implemented a field experiment with a 2x2 factorial design in a retail chain. We found evidence in line with the first two hypotheses that each practice itself raises a store's profit. However, we found no evidence for a complementarity between the two practices. Moreover, the bulk of the overall effect is generated through the provision of decision-facilitating information. The information intervention alone generated about €1,000 per month per store in profits with an approximated return on invest of 5,500%. When considering the net profit effect (subtracting the cost of the bonus) the combined intervention did not raise profits above the level attained by the information intervention alone.

Exploring survey data, we studied the underlying behavioral mechanisms in more detail. In particular, we found that the provision of information on profit margins led managers to focus their efforts much more on the placement of (higher margin) products – and this increase in the task focus occurred irrespective of whether store managers received performance pay or not. Moreover, detailed financial data shows an increase of sales revenues of high margin products as well as an increase of the number of high margin products sold. Hence, we find empirical support that our intervention indeed facilitated managers' decisions.

This study thus provides the first evidence from a field experiment within the same company which directly compares performance effects of the provision of decision facilitating information and influencing decisions through performance pay. Moreover, it provides evidence on the interplay of both practices. As the analysis shows, the provision of decision-facilitating information can have substantial effects on performance even in the absence of performance pay. Moreover, the additional introduction of performance pay may not raise performance substantially above the level achieved through information provision alone.

A key underlying question when comparing the two major roles of information in managerial accounting – namely the provision of decision-facilitating information and the use of decision-influencing information to set incentives – is the extent to which there is a-priori a conflict of interest between a firm and its managers. If interests are not well aligned, managers may not be motivated to effectively use the decision-facilitating information in their actions and, in turn, this information may be of little use. In this case, extrinsic incentives should be important to motivate managers to act on the information provided. If, however, managers have implicit incentives that align their interests with those of their employer, then the mere provision of decision-facilitating information can be very beneficial even when there are no direct explicit incentives. Indeed, as our empirical results show, substantial performance gains can be achieved through the provision of decision-facilitating information even in the absence of performance pay. We found that managers made use of this information and shifted their actions towards more profitable products and were able to raise profits accordingly.

3.9 References of Chapter 3

Anderson, Shannon W., and Amanda Kimball. “Evidence for the Feedback Role of Performance Measurement Systems.” *Management Science* 65.9 (2019): 3949-4450.

Athey, Susan, and Guido W. Imbens. “The Econometrics of Randomized Experiments.” in A. Benerjee and E. Duflo, *Handbook of Economic Field Experiments* 1 (2017): 73-140.

Bailey, Charles D., Lawrence D. Brown, and Antohny F. Cocco. “The Effects of Monetary Incentives on Worker Learning and Performance in an Assembly Task.” *Journal of Management Accounting Research* 10 (1998): 119-131.

Baiman, Stanley. “Agency research in managerial accounting: a survey.” *Journal of Accounting Literature* 1 (1982): 154-213.

Baker, George P. “Incentive Contracts and Performance Measurement.” *Journal of Political Economy* 100.3 (1992): 598-614.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Field experiments with firms." *Journal of Economic Perspectives* 25.3 (2011): 63-82.

Banker, Rajiv D., Seok-Young Lee, Gordon Potter, and Dhinu Srinivasan. “An Empirical Analysis of Continuing Improvements Following the Implementation of a Performance-Based Compensation Plan.” *Journal of Accounting and Economics* 30.3 (2000): 315-350.

Bassanini, Andrea, Alison L. Booth, Giorgio Brunello, Maria De Paola, and Edwin Leuven. “Workplace Training in Europe.” Education and Training in Europe (2007): Oxford University Press.

Bénabou, Roland, and Jean Tirole. “Incentives and Prosocial Behavior.” *American Economic Review* 96.5 (2006): 1652-1678.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. “Does Management Matter? Evidence from India.” *Quarterly Journal of Economics* 128.1 (2013): 1-51.

Bloom, Nicholas, James Liang, John Roberts, and Zhichung Jenny Ying. “Does Working from Home Work? Evidence from a Chinese Experiment.” *Quarterly Journal of Economics* 130.1 (2015): 165-217.

Bol, Jasmijn C., and Jeremy B. Lill. “Performance Target Revisions in Incentive Contracts: Do Information and Trust Reduce Ratcheting and the Ratchet Effect?” *The Accounting Review* 90.5 (2015): 1755-1778.

Bonner, Sarah E., Reid Hastie, Geoffrey B. Sprinkle, and S. Mark Young “A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting.” *Journal of Management Accounting Research* 12.1 (2000): 19-64.

Bonner, Sarah E., and Geoffrey B. Sprinkle. “The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research.” *Accounting, Organizations and Society* 27.4-5 (2002): 303-345.

Bouwens, Jan, and Laurence van Lent. “Assessing the Performance of Business Unit Managers.” *Journal of Accounting Research* 45.4 (2007): 667-697.

Bushman, Robert M., Raffi J. Indjejikian, and Mark C. Penno. "Private Predecision Information, Performance Measure Congruity, and the Value of Delegation." *Contemporary Accounting Research* 17.4 (2000): 561-587.

Casas-Arce Pablo, F. Asis Martínez-Jeres, and V. G. Narayanan. “The Impact of Forward-Looking Metrics on Employees Decision-Making: The Case of Customer Lifetime Value.” *The Accounting Review* 92.3 (2017a): 31-56.

Casas-Arce, Pablo, Sofia M. Lourenço, and F. Asis Martínez-Jeres. "The Performance Effect of Feedback Frequency and Detail: Evidence from a Field Experiment in Customer Satisfaction." *Journal of Accounting Research* 55.5 (2017b): 1051-1088.

Casas-Arce, Pablo, Martin Holzhacker, Matthias D. Mahlendorf, and Michal Matějka. ”Relative Performance Evaluation and the Ratchet Effect.” *Contemporary Accounting Research* 35.4 (2018): 1702-1731.

Chenhall, Robert H., and Frank Moers. “The Issue of Endogeneity within Theory-Based, Quantitative Management Accounting Research.” *European Accounting Review* 16.1 (2007): 173-196.

Dearden, Lorraine, Howard Reed, and John Van Reenen. “The Impact of Training on Productivity and Wages: Evidence from British Panel Data.” *Oxford Bulletin of Economics and Statistics* 68.4 (2006): 397-421.

Deci, Edward L. “Effects of externally mediated rewards on intrinsic motivation.” *Journal of Personality and Social Psychology 18.*1 (1971): 105-115.

De Grip, Andries, and Jan Sauermann. “The Effects of Training on own and co-worker productivity: Evidence from a field experiment.” *The Economic Journal* 122.560 (2012): 376-399.

Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke. “Tournament Incentives in the Field: Gender Differences in the Workplace.” *Journal of Labor Economics* 31.2 (2013): 305-326.

Demski, Joel S., and Gerald A. Feltham. “Cost determination: A conceptual approach.” Iowa State University Press (1976).

Evans, John H., Vicky B. Heiman-Hoffman, and Stephen E. Rau. “The accountability demand for information.” *Journal of Management Accounting Research* 6 (1994): 24.

Farrell, Anne M., Kathryn Kadous, Kristy L. Towry. “Contracting on Contemporaneous versus Forward-Looking Measures: An Experimental Investigation.” Contemporary Accounting Research 25.3 (2008): 773-802.

Feltham, Gerald A., and Jim Xie. "Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations." *The Accounting Review* 69.3 (1994): 429-453.

Frederickson, James R., Sean A. Peffer., and Jamie H. Pratt. “Performance Evaluation Judgments: Effects of Prior Experience under Different Performance Evaluation Schemes and Feedback Frequencies.” *Journal of Accounting Research* 37.1 (1999): 151-165.

Floyd, Eric, and John A. List. “Using Field Experiments in Accounting and Finance.” *Journal of Accounting Research* 54.2 (2016): 437-475.

Frey, Bruno S., and Reto Jegen. “Motivation Crowding Theory.” *Journal of Economic Surveys* 15.5 (2001): 589-611.

Friebel, Guido, Matthias Heinz, Miriam Krüger, and Nikolay Zubanov. “Team Incentives and Performance: Evidence from a Retail Chain.” *American Economic Review* 107.8 (2017): 2168-2203.

Ghosh, Dipankar. “De-escalation strategies: some experimental evidence.” *Behavioral Research in Accounting* 9 (1997): 88-111.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. "When and why incentives (don't) work to modify behavior." *Journal of Economic Perspectives* 25.4 (2011): 191-210.

Gosnell, Greer K., John A. List, and Robert Metcalfe. “The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains.” *Journal of Political Economy* (forthcoming).

Grafton, Jennifer, Anne M. Lillis, and Sally K. Widener. „The role of performance measurement and evaluation in building organizational capabilities and performance.” Accounting, Organizations and Society 35.7 (2010): 689-706.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. “Learning Through Noticing: Theory and Evidence from a Field Experiment.” *The Quarterly Journal of Economics* 129.3 (2014): 1311-1353.

Hemmer, T. “On the design and choice of ‘‘modern’’ management accounting measures.” *Journal of Management Accounting Research* 8 (1996): 87–116.

Hirshleifer, David, and Siew Hong Teoh. “Limited attention, information disclosure, and financial reporting.” *Journal of Accounting and Economics* 36.1-3 (2003): 337-386.

Hofmann, Christian, and Laurence Van Lent. “Organizational Design and Control Choices.” *Oxford Handbook of Strategy Implementation* (2017): 451-468.

Holderness, Kip, Kari Olsen, and Todd A. Thornock. “Assigned versus Chosen Relative Performance Information: The Effect of Feedback Frequency on Performance.” *Journal of Management Accounting Research* (2019): forthcoming.

Holmström, Bengt, and Paul Milgrom. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7 (1991): 24-52.

Holmström, Bengt, and Paul Milgrom. "The firm as an incentive system" *The American Economic Review* 84.4 (1994): 972-991.

Huffman, David, and Michael Bognanno. “High-Powered Performance Pay and Crowding Out of Nonmonetary Motives.” *Management Science* 64.10 (2018): 4471-4965.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi. “The effects of human resource management practices on productivity: A study of steel finishing lines.” *American Economic Review* 87.3 (1997): 291-313.

Indjejikian, Raffi J. “Performance Evaluation and Compensation Research: An Agency Perspective.” *Accounting Horizons* 13.2 (1999): 147-157.

Indjejikian, Raffi J., Michal Matějka, Kenneth A. Merchant, and Wim A. Van der Stede. “Earnings Targets and Annual Bonus Incentives.” *The Accounting Review* 89.4 (2014): 1227-1258.

Jensen, Michael C., and William H. Meckling. “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure.” *Journal of Financial Economics* 3.4 (1976): 305-360.

Jensen, Michael C., and Kevin J. Murphy. “Performance Pay and Top-Management Incentives.” *Journal of Political Economy* 2.98 (1990): 225-264.

Lazear, Edward P. “Performance Pay and Productivity.” *American Economic Review* 90.5 (2000): 1346-1361.

Lazear, Edward P. “Compensation and Incentives in the Workplace” *Journal of Economic Perspectives* 32.3 (2018): 195-214.

Lourenço, Sofia M. “Monetary Incentives, Feedback, and Recognition – Complements or Substitutes? Evidence from a Field Experiment in a Retail Service Company.” *The Accounting Review* 91.1 (2016): 279-297.

Mahlendorf, Matthias D., Michal Matejka, and Utz Schäffer. „Target Ratcheting and Achievability of Earnings Targets.” *Working Paper* (2015).

Manthei, Kathrin, Dirk Sliwka, and Timo Vogelsang. “Performance Pay and Prior Learning: Evidence from a Retail Chain.” *IZA Discussion Paper No. 11859* (2018).

Manthei, Kathrin, Dirk Sliwka, and Timo Vogelsang. “Talking about Performance or Paying for it?– Evidence from a Field Experiment.” *IZA Discussion Paper No. 12446* (2019).

Manthei, Kathrin, and Dirk Sliwka. “Multitasking and Subjective Performance Evaluations: Theory and Evidence from a Field Experiment in a Bank.” *Management Science* (forthcoming).

Merchant, Kenneth A. “Control in business organizations.” Boston, MA: Pitman Publishing, Inc. (1985).

Milgrom, Paul, and John Roberts. “Complementarities and fit strategy, structure, and organizational change in manufacturing.” *Journal of Accounting and Economics* 19.2-3 (1995): 179-208.

Mock, Theodore J. “Concepts of Information Value and Accounting.” *The Accounting Review* 46.4 (1971): 765-778.

Prendergast, Canice. “The Provision of Incentives in Firms.” *Journal of Economic Literature* 37.1 (1999): 7-63.

Presslee, Adam, Thomas W. Vance, and R. Alan Webb. “The Effects of Reward Type on Employee Goal Setting, Goal Commitment, and Performance.” *The Accounting Review* 88.5 (2013): 1805-1831.

Sandvik, Jason, Richard Saouma, Nathan Seegert, and Christopher Stanton. “Workplace Knowledge Flows.” *Working Paper* (2019).

Shearer, Bruce. “Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment.” *The Review of Economic Studies* 71.2 (2004): 513-534.

Simon, Herbert. A. “Centralization vs. decentralization in organizing the controller's department.” New York: Controllership Foundation (1954).

Sliwka, Dirk. “Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes.” *The American Economic Review* 97.3 (2007): 999-1012.

Sprinkle, Geoffrey B. “The Effect of Incentive Contracts on Learning and Performance.” *The Accounting Review* 75 (2000): 299-326.

Sprinkle, Geoffrey B. “Perspectives on experimental research in managerial accounting.” *Accounting, Organizations and Society* 28.2-3 (2003): 287-318.

Sunder, S. ”Theory of accounting and control.” Cincinnati, OH: South-Western Publishing Company (1997).

Van Lent, Laurence. “Endogeneity in Management Accounting Research: A Comment.” *European Accounting Review* 16.1 (2007): 197-205.

Van Veen-Dirks, Paula. “Different uses of performance measures: The evaluation versus reward of production managers.” *Accounting, Organizations and Society* 35.2 (2010) 141-164.

Wall, Friederike, and Dorothea Greiling. “Shareholder- versus Stakeholder-Orientation in Managerial Decision-Making.” *Review of Managerial Science* 5.2-3 (2011): 87-90.

3.10 Appendix to Chapter 3

3.10.1 Additional Tables and Figures

Table 3.9A – Classification of Store Manager Tasks

Task	Classification
Ordering of fruits and vegetables, plants Ordering of baked goods Ordering of meat Additional Ordering Baking of bakery articles	Ordering
Preparation of secondary placements Presentation and maintenance of special-offer tables (Non-Food/ Food/ end of aisle) Maintaining product positioning plans	Placements
Quality checks fruits, vegetables and plants Cleanliness of the baked goods stations Preservation and maintenance of the condition of the furnishings and the inventory (e.g., shelves, bumpers, freezers, cash desks) Guaranteeing the cleanliness and orderliness inside and outside the store	Cleanliness
Analysis of Spoilage Analysis of Sales Analysis of Personnel Costs Analysis of Hourly Output Analysis of Inventory	KPI
Checking minimum durability date (meat, dairy, convenience) Process left overs Stocking of goods and maintenance of shelves (colonial goods, frozen goods, load) Incoming goods inspection Security of goods Working on gap listing and inventory care	Inventory
Training of cashier employees Appraisal interviews / leadership Staff planning	Personnel Management
Communication with customers, processing of customer requests Own cashier work (Temporary price reductions)	Own Effort

Table 3.10A – Balancing Table

	(1) Descriptives Overall	(2) Descriptives Control	(3) Descriptives Information	(4) Descriptives Bonus	(5) Descriptives Bonus&Information
Profits Jan-Mar '17	34244.12	32535.37	33261.79	35116.27	36102.65
	(14444.84)	(13805.71)	(12525.05)	(15051.03)	(16141.77)
Planned Profits Jan-Mar '17	34437.85	32880.53	33586.91	35642.64	35690.4
	(13635.98)	(12873.05)	(11890.28)	(15244.92)	(14363.22)
Female Store Manager (Y/N)	0.68	0.76	0.65	0.60**	0.72
	(0.47)	(0.43)	(0.48)	(0.49)	(0.45)
Walking Customers (Y/N)	0.12	0.10	0.18	0.15	0.05
	(0.33)	(0.30)	(0.39)	(0.36)	(0.23)
FTE	6.63	6.45	6.69	6.84*	6.55
	(1.38)	(1.17)	(1.39)	(1.58)	(1.33)
Age of Store	16.37	17.63	16.57	17.44	13.89**
	(9.75)	(10.47)	(1.39)	(10.10)	(8.11)
Age Store Manager	43.15	44.57	43.52	41.15**	43.25
	(10.84)	(10.05)	(10.55)	(10.79)	(10.99)
Tenure Store Manager	14.18	15.51	14.23	13.01**	13.96
	(8.44)	(8.43)	(8.64)	(7.73)	(8.82)
Store Space	695.89	701.70	679.03	693.33	709.45
	(134.09)	(112.95)	(143.24)	(121.67)	(154.03)
N Observations	363	91	92	88	92

Note: The table reports means of the respective variables for the different treatment groups and their standard deviations in parentheses. Asterisks display significance levels from t-tests (fisher exact test for binary variables) of the respective treatment group against the control group. * p<0.1, ** p<0.05, *** p<0.001.

Table 3.11A – Regressions only using Treatment Period

	(1)	(2)	(3)	(4)
	OLS	log OLS	OLS	log OLS
Treatment Effect	411.41	0.0162	507.62	0.0206*
BONUS	(342.63)	(0.0104)	(357.36)	(0.0106)
Treatment Effect	721.29**	0.0149	706.15**	0.0297***
INFORMATION	(336.68)	(0.0111)	(330.44)	(0.0109)
Treatment Effect	907.57**	0.0221**	734.83*	0.0238**
BONUS & INFORMATION	(364.29)	(0.0106)	(391.75)	(0.0105)
Time FE	Yes	Yes	Yes	Yes
Store FE	No	No	No	No
District Manager FE	No	No	No	No
Store Manager FE	No	No	No	No
Refurbishments	Yes	Yes	Yes	Yes
Planned Profits	Yes	Yes	Yes	Yes
Further Controls	No	No	Yes	Yes
N Observations	1086	1086	1068	1068
N Stores	363	363	356	356
Cluster	56	56	56	56
Within R^2				
Overall R^2	0.9260	0.9129	0.9278	0.9082

Note: The table reports results from ordinary least squares estimations with profits at the store level as the dependent variable in columns 1&3 and the log value in columns 2&4. Regressions control for the mean of profits from January 2016 - March 2017 and the randomization pair. All regressions further control for possible refurbishments of a store and the companies planed profits. Columns 3&4 further control for variables with slight imbalance between treatments (gender, FTE, age of the store, age of the store manager, tenure of the store manager). Observations are excluded once a store manager switched the store during the treatment period. Robust standard errors are clustered on the district level at the start of the experiment and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.001.

Table 3.12A – Monthly Treatment Effects

		Full sample (ITT)		
	(1)	(2)	(3)	(4)
	Profits	Profits	Ln (Profits)	Ln (Profits)
Treatment Effect	1315.6***	1742.5***	0.0284**	0.0397***
BONUS 1st Month	(473.4)	(516.8)	(0.0138)	(0.0143)
Treatment Effect	683.0	1202.5**	0.0212	0.0348**
BONUS 2nd Month	(533.7)	(556.0)	(0.0148)	(0.0150)
Treatment Effect	-254.9	186.0	-0.00557	0.00635
BONUS 3rd Month	(823.3)	(901.9)	(0.0214)	(0.0232)
Treatment Effect	-613.3	-202.4	-0.0117	-0.00152
BONUS 4th Month (after experiment)	(907.1)	(937.8)	(0.0225)	(0.0231)
Treatment Effect	-1821.0	-1329.5	-0.0357	-0.0192
BONUS 5th Month (after experiment)	(1227.8)	(1324.1)	(0.0331)	(0.0353)
Treatment Effect	1568.2***	1755.0***	0.0343*	0.0396**
INFORMATION 1st Month	(569.5)	(594.6)	(0.0182)	(0.0191)
Treatment Effect	1059.2*	1444.9**	0.0180	0.0276
INFORMATION 2nd Month	(570.6)	(588.9)	(0.0189)	(0.0205)
Treatment Effect	490.1	853.5	0.00339	0.0128
INFORMATION 3rd Month	(832.8)	(933.1)	(0.0225)	(0.0249)
Treatment Effect	693.7	1028.7	0.0162	0.0242
INFORMATION 4th Month (after experiment)	(896.2)	(924.3)	(0.0241)	(0.0252)
Treatment Effect	-107.9	277.5	-0.00044	0.00962
INFORMATION 5th Month (after experiment)	(954.2)	(1008.2)	(0.0284)	(0.0296)
Treatment Effect	1488.0***	1492.5***	0.0319**	0.0325**
BONUS & INFORMATION 1st Month	(535.4)	(544.9)	(0.0142)	(0.0152)
Treatment Effect	1681.6**	1958.2**	0.0349*	0.0404*
BONUS & INFORMATION 2nd Month	(749.0)	(780.0)	(0.0195)	(0.0212)
Treatment Effect	908.0	1208.9	0.0142	0.0209
BONUS & INFORMATION 3rd Month	(815.8)	(923.7)	(0.0214)	(0.0241)
Treatment Effect	-1104.7	-552.4	-0.00323	-0.001
BONUS & INFORMATION 4th Month (after experiment)	(1143.5)	(1063.5)	(0.0243)	(0.0246)
Treatment Effect	-752.2	-428.9	-0.00567	0.00238
BONUS & INFORMATION 5th Month (after experiment)	(1120.7)	(1220.8)	(0.0343)	(0.0368)
Fixed Effects (Store, Time)	Yes	Yes	Yes	Yes
Fixed Effects (Store Manager, District Manager)	No	Yes	No	Yes
Refurbishments	Yes	Yes	Yes	Yes
Planned Profits	Yes	Yes	Yes	Yes
N Observations	7196	7007	7193	7006
N Store	363	363	363	363
Cluster	56	56	56	56
Within R^2	0.3351	0.3553	0.3591	0.3469
Overall R^2	0.8285	0.7379	0.7948	0.7276

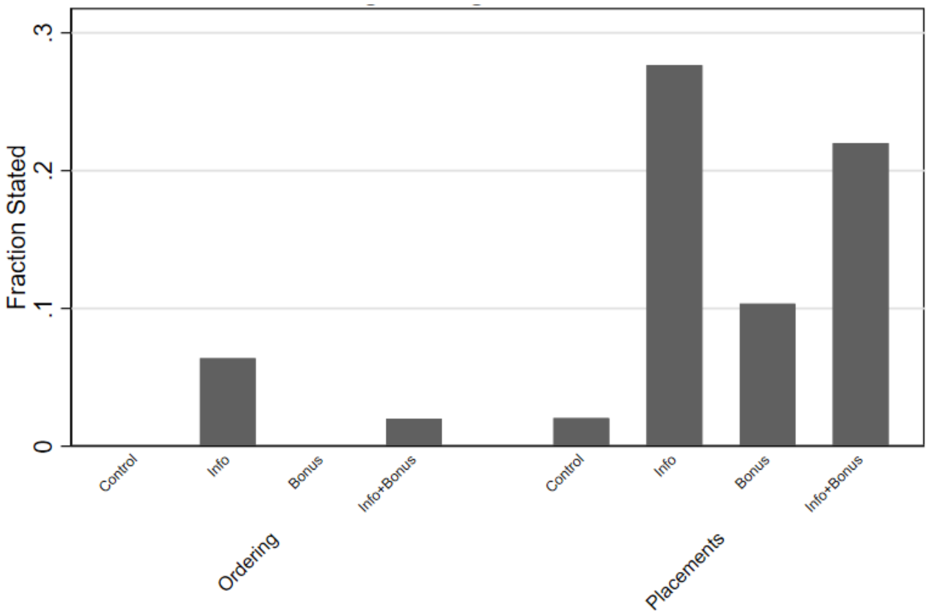
Note: The table reports results from fixed effects regressions with the profits on the store level as the dependent variable. The regression accounts for time and store fixed effects (columns 1-8) and adds fixed effects for district and store managers in columns 2,4,6,8. Columns 5-8 drop observations for store managers who did not watch the training video during the treatment time. The fixed effects regressions compare pre-treatment observations (January 2016 - March 2017) with the observations during the experiment (April 2017 - June 2017). All regressions control for possible refurbishments of a store and the companies planned value. Observations are excluded once a store manager switched the store during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.001.

Table 3.13A – Self-Stated Work Focus and 2nd Questionnaire (Closed Questions)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A	Ordering	Placements	Cleanliness	KPI	Inventory	Personnel Management	Own Effort
Treatment Effect INFORMATION	0.0650 (0.127)	0.0873 (0.207)	0.269* (0.153)	0.267 (0.207)	0.155 (0.138)	0.273 (0.166)	0.169 (0.187)
Treatment Effect BONUS	-0.115 (0.120)	-0.172 (0.197)	0.0924 (0.146)	0.191 (0.196)	-0.0457 (0.131)	-0.0425 (0.157)	-0.259 (0.178)
Treatment Effect INFORMATION &BONUS	0.0354 (0.125)	0.0556 (0.204)	0.144 (0.151)	0.205 (0.203)	0.154 (0.136)	0.118 (0.163)	0.136 (0.184)
N Observations	204	204	204	204	204	204	204
R^2	0.013	0.010	0.016	0.009	0.018	0.022	0.035
Panel B	Ordering	Placements	Cleanliness	KPI	Inventory	Personnel Management	Own Effort
Treatment Effect INFORMATION	0.0152 (0.123)	0.0930 (0.214)	0.227 (0.147)	0.257 (0.206)	0.146 (0.140)	0.337** (0.169)	0.206 (0.175)
Treatment Effect BONUS	0.0307 (0.124)	0.0430 (0.216)	0.207 (0.149)	0.468** (0.209)	0.0536 (0.142)	0.162 (0.171)	0.0359 (0.177)
Treatment Effect INFORMATION &BONUS	0.0372 (0.121)	0.0881 (0.210)	0.140 (0.144)	0.164 (0.202)	0.126 (0.138)	0.124 (0.165)	0.0948 (0.172)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N Observations	181	181	181	181	181	181	181
R^2	0.100	0.121	0.185	0.119	0.096	0.121	0.191

Note: The table reports results from OLS regressions with the stated amount of different tasks to increase profits from an online questionnaire as the dependent variable. The variables range from 1 (low focus) to 6 (high focus). Panel B control include the size of the store, amount of full time equivalent employees (FTE), age of the store manager, and the annual subjective performance evaluation. Observations are excluded once a store manager switched the store during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * p<0.1, ** p<0.05, *** p<0.001.

Figure 3.5A – Focus on High Margin Products



Note: The figure displays the fraction of explicitly stated tasks dimensions with a focus on high margin products to increase profits obtained from open questions of an ex-post questionnaire. N=204.

3.10.2 Instructions

3.10.2.1 Store Manager - INFORMATION Group (send to their home address, originally in German)

Project DB1[27]

Dear Mr./Mrs XXX,

a positive DB1 profit measure is important for the economic success of [*the company*]. For this reason, the DB1 project will be implemented in your region during the next few months. Within the scope of the DB1 project, you will have the opportunity to receive a learning unit.

You will now have access to the learning unit.

Learning Unit[28]:

In order to renew and deepen your knowledge about the DB1, we have put together an online learning unit for you. This consists of a short learning video and a quiz afterwards. In order for [*the company*] to remain economically strong, you should finish this learning unit by 08.04.2017!

The learning unit is provided by the *[name of university]*. You can complete the learning unit using the access data listed below in the EDP (Home left> Section "Other"), with your private computer or your smartphone. Please see the access data listed below.

Access data learning unit:
Please visit the following website for the learning unit:

Your password is:

Alternatively, you can also use the following QR code directly:

In order for you to keep track of the explained figures, you will receive a separate DB1-report in the Store Data Warehouse at the end of the following month.

We would like to thank you sincerely in advance for your participation and support.

If you have any questions, please contact your district management / personnel management.

Yours sincerely,

[27] The company uses „DB1" (short for Deckungsbeitrag 1/ contribution margin) as an internal title for the simplified profit measure explained above in our study: *Profit = Net Sales – Cost of Goods Sold – Staff Costs – Inventory Losses.*

[28] Due to previous company wording, the company uses "learning unit" as an internal description for the learning video, the quiz, the margin information and the monthly feedback. We refer to this as "information package" in the above.

3.10.2.2 Store Manager - BONUS Group (send to their home address, originally in German)

Dear Mr./Mrs XXX,

a positive DB1 profit measure is important for the economic success of [*the company*]. For this reason, the DB1 project will be implemented in your region during the next few months. Within the scope of the DB1 project, you will have the opportunity to earn an additional bonus

Your bonus period starts on 01.04.2017 for 3 months.

<u>Bonus:</u>

Within this project, you will be able to earn an additional bonus in your store over the next three months (April, May, June) for increasing the DB1 profit measure.

Therefore, the DB1 profit measure of your store will be compared monthly with the plan DB1 of the respective month. If your DB1 profit measure is more than 80% of the plan DB1, you will receive a bonus. From the difference between the DB1 profit measure and 80% of the plan DB1, you are paid-out 5% as a premium in euros.

Calculation: DB1-Bonus (in €) = (DB1 – 80% of the Plan DB1) * 0.05

The DB1-Bonus is always calculated at the end of the month. The sum of the bonuses from the three months will be paid out to you in September 2017 with your payroll. This means that the bonus amount can be negative in a single month (if the plan achievement is under 80%). Should you still have a negative amount after the end of the three months, you will be paid € 0. Please see the attached info sheet for the bonus calculation.

Information about your bonus amount will always be sent by post to your home at the end of the following month.

We would like to thank you sincerely in advance for your participation and support.

If you have any questions, please contact your district management / personnel management.

Yours sincerely,

3.10.2.3 Store Manager – BONUS&INFORMATION Group (send to their home address, originally in German)

Dear Mr./Mrs XXX,

a positive DB1 profit measure is important for the economic success of [*the company*]. For this reason, the DB1 project will be implemented in your region during the next few months. Within the scope of the DB1 project, you will have the opportunity to earn an additional bonus, and receive a learning unit.

Your bonus period starts on 01.04.2017 for 3 months. You will now have access to the learning unit.

<u>Bonus:</u>

Within this project, you will be able to earn an additional bonus in your store over the next three months (April, May, June) for increasing the DB1 profit measure.

Therefore, the DB1 profit measure of your store will be compared monthly with the plan DB1 of the respective month. If your DB1 profit measure is more than 80% of the plan DB1, you will receive a bonus. From the difference between the DB1 profit measure and 80% of the plan DB1, you are paid-out 5% as a premium in euros.

Calculation: DB1-Bonus (in €) = (DB1 – 80% of the Plan DB1) * 0.05

The DB1-Bonus is always calculated at the end of the month. The sum of the bonuses from the three months will be paid out to you in September 2017 with your payroll. This means that the bonus amount can be negative in a single month (if the plan achievement is under 80%). Should you still have a negative amount after the end of the three months, you will be paid € 0. Please see the attached info sheet for the bonus calculation.

Information about your bonus amount will always be sent by post to your home at the end of the following month.

<u>Learning Unit:</u>

In order to renew and deepen your knowledge about the DB1, we have put together an online learning unit for you. This consists of a short learning video and a quiz afterwards. In order for [*the company*] to remain economically strong, you should finish this learning unit by 08.04.2017!

The learning unit is provided by the *[name of university]*. You can complete the learning unit using the access data listed below in the EDP (Home left> Section "Other"), with your private computer or your smartphone. Please see the access data listed below.

<u>Access data learning unit:</u>

Please visit the following website for the learning unit:

Your password is:

Alternatively, you can also use the following QR code directly:

In order for you to keep track of the explained figures, you will receive a separate DB1-report in the Store Data Warehouse at the end of the following month.

We would like to thank you sincerely in advance for your participation and support.

If you have any questions, please contact your district management / personnel management.

Yours sincerely,

3.10.2.4 Monthly Communication to Store Manager (send to their home address, originally in German)

Project DB1

Dear Mr./Mrs XXX,

Please find below a summary of your key figures in the first month of the project.

Summary of your DB1 profit measure in April 2017:[29]
(Amounts are not rounded until the end)

Sales:
Cost of goods sold:
Personnel costs:
Inventory losses:

This results in a DB1 April/2017:
For a plan DB1 April/2017:

The resulting premium amount for the month of April is:

(DB1 – 0.8 * plan DB1) * 0.05 € =

Summary of your bonus amounts since April 2017:
(Amounts are not rounded until the end)

Bonus amount April 2017: € (gross)

The sum of the bonus amounts (if greater than 0) will be paid-out at the end of the three-month period in September 2017 with your payroll. Please note that positive bonus amounts are offset against negative ones. There will only be one bonus payment of the grand total in September.

For further questions, please contact your district manager / personnel management.

[29] For accounting reasons, the letter in May came with an additional information: "In April, adjusting entries through accounting were posted to the region only and not distributed to the branches. Their profit margin is therefore too well represented. These bookings will be made up with the May-finalization. Therefore, you will find the margin correction in your May letter with a reversed sign. In sum of April and May, the correction value will be € 0.00. We ask for your understanding."

3.10.2.5 District Manager – CONTROL Group (sent to their e-mail address, originally in German)

Dear Mr./Mrs XXX.,

A positive DB1 profit measure is important for the economic success of [*the company*]. For this reason, the DB 1 project will be implemented in your region during the next few months.

Within the scope of the DB 1 project, all store managers will receive a learning unit about DB 1 in the near future. Moreover, an additional bonus for store managers is introduced in all stores of the region. For administrative reasons, the bonus will be introduced in the districts at different times. The assignment happens randomly according to a statistical procedure.

In your district, store managers will receive the bonus and the learning unit at a <u>later date</u>. You will be informed in sufficient time about the exact time frame.

<u>Communication upon inquiries of store managers</u>:

- If your store managers ask why they did not receive the learning unit about the DB1 profit measure yet, we ask that you communicate that this is a random selection and that your area's store managers will receive the learning unit at a later time.
- If your store managers ask why they are not getting a bonus for the increased DB1 profit measure, we also ask you to communicate that this is a random selection and that the store managers in your district will in any case receive a bonus at a later date.

For a neat evaluation, it is important that all district managers strictly follow this language regulation. Please do not pass any further information on to store managers and only discuss the learning unit and the bonus if a store manager explicitly asks for it.

The findings of this project are of great importance to [*the company*].

For inquiries your personnel management is at your disposal at any time.

Yours sincerely

3.10.2.6 District Manager – INFORMATION Group (sent to their e-mail address, originally in German)

Dear Mr./Mrs XXX.,

A positive DB1 profit measure is important for the economic success of *[the company]*. For this reason, the DB 1 project will be implemented in your region during the next few months.

Within the scope of the DB 1 project, all store managers will receive a learning unit about DB 1 in the near future. Moreover, an additional bonus for store managers is introduced in all stores of the region. For administrative reasons, the bonus will be introduced in the districts at different times. The assignment happens randomly according to a statistical procedure.

In your district, store managers will receive the bonus at a later date. You will be informed in sufficient time about the exact time frame.
However, store managers will have access to an online learning unit with regard to the contribution margin 1 from 27.03.2017. Please make sure that the learning unit is completed by the store managers in your district.

Your store managers will be informed elaborately and separately by mail.

<u>Store manager learning unit:</u>

In order to renew and deepen the knowledge of store managers regarding the DB1 profit measure, we have put together an online learning unit for your store managers. This consists of a learning video and a quiz afterwards. If you are interested, you can also watch the learning video (provided by the *[name of university]*) with the following link: XXXX
Your personal password is: XXXXX

<u>Communication upon inquiries of store managers:</u>

If your store managers ask why they are not getting a bonus for the increased DB1 profit measure, we also ask you to communicate that this is a random selection and that the store managers in your district will in any case receive a bonus at a later date.

For a neat evaluation, it is important that all district managers strictly follow this language regulation. Please do not pass any further information on to store managers and only discuss the learning unit and the bonus if a store manager explicitly asks for it.

The findings of this project are of great importance to [*the company*].

For inquiries your personnel management is at your disposal at any time.

Yours sincerely

3.10.2.7 District Manager – BONUS Group (sent to their e-mail address, originally in German)

Dear Mr./Mrs XXX.,

A positive DB1 profit measure is important for the economic success of *[the company]*. For this reason, the DB 1 project will be implemented in your region during the next few months.

Within the scope of the DB 1 project, all store managers will receive a learning unit about DB 1 in the near future. Moreover, an additional bonus for store managers is introduced in all stores of the region. For administrative reasons, the bonus will be introduced in the districts at different times. The assignment happens randomly according to a statistical procedure.

In your district, store managers will receive a bonus from 01.04.2017 for 3 months.
However, they will have only access to the online learning unit about the contribution margin 1 at a later time.

Your store managers will be informed elaborately and separately by mail.

<u>Store manager bonus:</u>

For the bonus of your store managers, the DB! Profit measure of the respective store will be compared monthly with the planned DB1 of the respective month. If the DB1 is more than 80% of the planned DB1, the store manager will receive a bonus. From the difference between the DB1 and 80% of the planned DB1, the store manager is paid-out 5% as a bonus in euros.

Calculation: DB1-Bonus (in €) = (DB1 – 80% of the plan DB1) * 0.05

The bonus is always calculated at the end of the month. The sum of the bonuses from the three months will be paid-out to store managers in September 2017 with their payroll. This means that the store manager can have a negative bonus amount in a single month (if the plan achievement is under 80%). Should they still have a negative amount after the end of the three months, they will be paid € 0. In addition, your store managers will receive a monthly report on the development of their DB1 (accessible in the Store Data Warehouse) and their bonus amounts (by mail home).

<u>Communication upon inquiries of store managers:</u>

If your store managers asks why they do not receive the learning unit about DB1 profit measure yet, we ask that you communicate that this is a random selection and that your area's store managers will receive the learning unit at a later time.

For a neat evaluation, it is important that all district managers strictly follow this language regulation. Please do not pass any further information on to store managers and only discuss the learning unit and the bonus if a store manager explicitly asks for it.

The findings of this project are of great importance to [*the company*].

For inquiries your personnel management is at your disposal at any time.

Yours sincerely,

Your store managers will be informed separately with a letter. This letter also includes the following information about the DB1-Bonus.

3.10.2.8 District Manager – BONUS&INFORMATION Group (sent to their e-mail address, originally in German)

Dear Mr./Mrs XXX.,

A positive DB1 profit measure is important for the economic success of *[the company]*. For this reason, the DB 1 project will be implemented in your region during the next few months.

Within the scope of the DB 1 project, all store managers will receive a learning unit about DB 1 in the near future. Moreover, an additional bonus for store managers is introduced in all stores of the region. For administrative reasons, the bonus will be introduced in the districts at different times. The assignment happens randomly according to a statistical procedure.

In your district, store managers will receive a bonus from 01.04.2017 for 3 months. Store managers will also have access to an online learning unit regarding the contribution margin 1 from 27.03.2017. Please make sure that the learning unit is completed by the store managers in your district.

Your store managers will be informed elaborately and separately by mail.

<u>Store manager learning unit:</u>

In order to renew and deepen the knowledge of store managers regarding the DB1 profit measure, we have put together an online learning unit for your store managers. This consists of a learning video and a quiz afterwards. If you are interested, you can also watch the learning video (provided by the *[name of university]*) with the following link: XXXX
Your personal password is: XXXXX

<u>Store manager bonus:</u>

For the bonus of your store managers, the DB! Profit measure of the respective store will be compared monthly with the planned DB1 of the respective month. If the DB1 is more than 80% of the planned DB1, the store manager will receive a bonus. From the difference between the DB1 and 80% of the planned DB1, the store manager is paid-out 5% as a bonus in euros.

Calculation: DB1-Bonus (in €) = (DB1 – 80% of the plan DB1) * 0.05

The bonus is always calculated at the end of the month. The sum of the bonuses from the three months will be paid-out to store managers in September 2017 with their payroll. This means that the store manager can have a negative bonus amount in a single month (if the plan achievement is under 80%). Should they still have a negative amount after the end of the three months, they will be paid € 0. In addition, your store managers will receive a monthly report on the development of their DB1 (accessible in the Store Data Warehouse) and their bonus amounts (by mail home).

For a neat evaluation, it is important that all district managers strictly follow this language regulation. Please do not pass any further information on to store managers and only discuss the learning unit and the bonus if a store manager explicitly asks for it.

The findings of this project are of great importance to [*the company*].

For inquiries your personnel management is at your disposal at any time.

Yours sincerely,

Your store managers will be informed separately with a letter. This letter also includes the following information about the DB1-Bonus.

3.10.2.9 Information on the bonus calculation (attached to all letters of store and district managers in the bonus treatments)

<u>Information about the DB1-Bonus</u>

The DB1 profit measure represents the economic success of [*the company*]. The more positive it is, the stronger [*the company*] is positioned. The DB1 profit measure is the net sales minus costs that can be influenced such as Inventory and personnel costs.

Please find attached the details for the calculation as well as a fictitious example.

<u>Calculation DB1-Bonus</u>

From 01.04.2017 up to and including 30.06.2017, you will be informed monthly about the increase of your DB1 profit measure compared to your plan of the DB1.

If your DB1 profit measure is at least 80% of the plan DB1, you will receive a bonus. From the difference between your actual DB1 profit measure and 80% of the plan DB1, you are paid-out 5% as a bonus in euros.

Amount in euros = (DB1 – 80% plan DB1) * 0.05

This amount in euros is added up for the months of April, May and June and then paid out to you with your payroll in September.

<u>Fictious Example</u>

Month April: The DB1 in April was 30.000 with a plan DB1 of 28.000.

This results in a euro amount of (30000 – 0.8 * 28000) * 0.05 = 380 Euro.

Month May: The DB1 in April was 24.000 with a plan DB1 of 29.000.

This results in a euro amount of (22000 – 0.8 * 29000) * 0.05 = - 60 Euro.

Month June: The DB1 in April was 28.000 with a plan DB1 of 29.000.

This results in a euro amount of (28000 – 0.8 * 29000) * 0.05 = 240 Euro.

Total bonus paid: 380 (April) – 60 (May) + 240 (June) = 560€

Thus, in September 560 € would be paid as a bonus.

3.10.3 The Learning Video Excerpts of the Script

Opening Scene:

You know the situation: you have to list new articles and remove goods from the assortment every month. But according to which criteria does this selection take place?

An important goal in this process is to increase the DB1.

The variety of business key figures can quickly cause confusion. What is actually behind the parameters of net sales, margin and contribution margin and what is the difference between them?

To renew your knowledge regarding the DB1, we have put together this learning unit for you.

Excerpt 2:

As a store manager, you are responsible, among other things, to place products optimally in the store, to order goods in the right quantity at the right time and to plan the deployment of staff efficiently in your market.

We would like to illustrate with some examples how you can directly influence the DB1 through your activities in your store.

The “DB1” can be influenced by four central levers.

Net sales in combination with margin, inventory discrepancies and losses as well as personnel costs.

Excerpt 3:

For better orientation, we have developed a system. We have sorted all products by their margin and divided them into five categories. The 20% of the products with the highest margin will be marked SP1 in your PDE device in the product information. The 20% of the lowest-margin products will be marked SP5 in your PDE device. Accordingly, SP2, SP3 and SP4 identify products in between. Here we use the school grade logic.

Note, however, that of course, the SP5 products should not be neglected, because they can be good for e.g. customer acquisition.

Chapter 4

4. Time is (not) Money – Incentive Effects of Granting Leisure Time

4.1 Introduction

Flexible work arrangements are becoming an important characteristic of modern labor markets.[1] In light of this trend, the US government implemented a novel bonus that awards employees with additional paid leisure time at home (time-off award).[2] As this potentially decreases the marginal utility of leisure, it may also address the growing problem of on-the-job leisure consumption, which costs companies billions of dollars because it decreases efficient working time (see, e.g., Malachowski 2005, D'Abate and Eddy 2007, Conner 2015, Park 2019). In this real-effort experiment, I show that granting leisure time as a gift increases performance through the mechanism of less on-the-job leisure (Internet) consumption as distortionary behavior influencing the intensive margin of labor supply. To provide some external validity, I conduct an online survey experiment in which human resource managers anticipate this mechanism and recognize even more advantages of granting additional leisure time.

The question of what motivates performance is central to business and economics (Prendergast 1999, Bonner and Sprinkle 2002, DellaVigna and Pope 2018). Monetary short-term bonuses, often unconditional on future effort, are a common management practice to motivate employees to increase performance (WorldatWork survey 2016). The design of such bonuses range from performance-contingent incentives to awards, rewards, and gifts.[3] As claimed by Akerlof (1982), employees reciprocate such above-market-clearing wages with higher performance. This is especially important if a worker's compensation is solely based on the amount of hours they spend at the workplace, with possible distortionary behavior by workers influencing the effective hours worked.

[1] See for instance Bloom et al. (2015), Mas and Pallais (2017), Wiswall and Zafar (2018), Katz and Krueger (2019), and Sherman (2019).

[2] Leisure seems to be not only more valued nowadays compared to previous generations (Twenge et al. 2010, Wiswall and Zafar 2018), but it is also becoming more prominent as a wage substitute. German railway employees, for instance, could recently decide between a wage increase and more time off. More than half of the employees decided for more time off (*Handelsblatt Global*, December 14, 2016). The German labor union IG Metal let German employees decide whether they want to have a wage increase or more paid vacation. Among shift workers, for instance, 70-80% chose to have more paid vacation (*Handelsblatt Global*, November 11, 2018).

[3] I will refer in the following to a "reward" as something given in exchange for something, a "gift" as something given voluntarily without any expectations, and an "award" as something one could earn, for example in a competition. A "bonus" will refer to all three of those forms of additional compensation.

However, the empirical (causal) evidence from laboratory and field experiments is mixed, with unconditional monetary bonuses often having close to zero average treatment effects on performance.[4] Non-monetary bonus domains seem to be more consistent in successfully increasing employees' performance. Awarding performance with symbolic recognition, for instance, increases performances significantly (Neckermann and Kosfeld 2011, Ashraf et al. 2014, Bradler et al. 2016, Lourenço 2016, Gallus 2017).[5] Providing employees with a non-monetary gift in the form of a thermos bottle before they start working, Kube et al. (2012) find a 25% performance increase compared to the baseline of no gift. Interestingly, a monetary gift only increases performance by 5%, which is not statistically significant. Bradler and Neckermann (2019) also study monetary gifts and recognition, and find a similar but significant performance effect of 4-5% in both treatments. Importantly, this literature still does not cover all of the possible non-monetary bonus domains that can be observed in practice and, thus, remains uncertain about the specific performance effects of different non-monetary reward domains (Gibbs 2017, Merchant and Van der Stede 2017).

This paper takes the important path of further investigating non-standard bonuses. Using leisure time as a bonus domain, it contributes to the literature a novel non-monetary domain with notable characteristics. A possible theoretical effect would be that, assuming concave utility of overall leisure, the increase of leisure time at home reduces the marginal utility of on-the-job leisure and, thus, its consumption. This hypothesized reduction in on-the-job leisure would lead to a more effective labor supply during working hours (intensive margin of labor supply). From the point of the classical economic literature (e.g. Heckman 1993, Prescott 2004), on-the-job leisure consumption might be a novel domain of the intensive margin of labor supply. Usually studies use weekly hours worked (the availability of an employee) as a proxy for the intensive margin. This can neglect the different margins an employee can react on, as employees do not have to be productive while present and, for instance, can consume leisure time.[6] Moreover, studies on the relationship between productivity and hours worked could be

[4] Laboratory experiments in which subjects choose/state an effort level from a convex cost of effort set find that subjects reciprocate an unconditional monetary payment or a higher wage (see, e.g., Fehr et al. 1993, Hannan et al. 2002, Charness 2004). Laboratory experiments using a real-effort task, however, find close to zero average treatment effects (Englmaier and Leider 2012, Carpenter 2016, Neckermann and Yang 2017). An exception here is Sliwka and Werner (2017). However, they employ a dynamic work setting in which the unconditional monetary payment (wage) changes over eight rounds. In line with the other studies, effort in their first round does not significantly vary between different wage levels. Performance increases are driven by wage increases in later rounds. Field experiments also find mixed results (positive effects are reported in Gneezy and List 2006, Falk 2007, Kube et al. 2012, and Gilchrist et al. 2016. No or little evidence of gift exchange in the field is demonstrated in Henning-Schmidt et al. 2010, DellaVigna et al. 2016, Esteves-Sorenson 2018, and DellaVigna and Pope 2018. Kube et al. 2013 and Cohn et al. 2014 find an effect in case of wage cuts but no effect for wage increases).

[5] An exception here might be Charness et al. (2014) who do not find a statistically significant performance difference for laboratory subjects who receive a gold medal in treatments with performance rankings.

[6] Few empirical studies exist on shirking/loafing/slacking off at the workplace. These studies, however, mainly focus on absenteeism (extensive margin of labor supply) (see, e.g., Ichino and Maggi 2000), on individual productivity decreases (Park 2019) or rushing through the job (Chan 2018) instead of eliciting distortionary behavior itself during work.

confounded by the possibility that the definition of hours worked still leaves room for unused working hours.

Although non-monetary bonuses often work better in increasing employees' performance, they are missing two important and typical characteristics compared to monetary bonuses. First, employees cannot make unconstrained consumption choices (fungibility), and, second, the non-monetary bonus often cannot be paid out in marginal units (divisibility). The time domain shares these two favorable characteristics with monetary bonuses, while maintaining the favorable characteristics of other non-monetary bonuses. Additionally, time is a scarce resource similar to money, and it is subject to very frequent and often unconscious decisions (Leclerc et al. 1995). Importantly, time as a bonus domain comes at no additional cost to the employer, other than the loss of output from employees leaving work earlier. This, however, may be (partially) offset if employees reduce their on-the-job leisure time, and the firm potentially saves on costs associated with longer working hours. Using leisure time as an incentive may additionally also generate positive spillovers for instance on employees' happiness (e.g. Whillans et al. 2017), felt appreciation (e.g. Wiswall and Zafar 2018), health (e.g. Bannai and Tamakoshi 2014), and private relationships (e.g. Sherman 2019). Employees might then further reciprocate this with an increased productivity (see, for instance, Oswald et al. 2015, Bradler and Neckermann 2019). Despite these interesting features, the research on time as a reward or decision medium in experiments is still small and mainly focused on social preference (Noussair and Stoop 2015, Danilov and Vogelsang 2016, Brown et al. 2019) and decisions under risk (Festjens et al. 2015).

This paper studies leisure time as a bonus domain, using a laboratory setting to isolate incentive effects and mechanisms of time bonuses from other external forces present in corporate environments (e.g., reputational concerns, career concerns, peer effects, supervisor monitoring). Moreover, performance and on-the-job leisure consumption can be precisely measured in the laboratory. The experiment is set up as follows. Initially, subjects in the experiment face two real-effort working periods of 30 minutes each. They also have the possibility to consume on-the-job leisure (Internet) at any time. Subjects receive an unconditional base compensation for both working periods. Treatments then vary the compensation domain (leisure time at home or money) of the gift, which is unconditional on later performance. Importantly, the gift is framed as an additional compensation and announced before the first working period.[7] Subjects receive the gift for the first period, and it is common

[7] Please note, that in contrast to previous studies no appreciation of the work occurs (as for instance in Kube et al. 2012). Throughout the experiment, I use the value-free wording of a "base compensation" and an "additional compensation" and a

knowledge that there will be no gift in the second one. With a gift of leisure time, subjects leave the laboratory earlier during the second working period.

Confirming the recent literature on unconditional monetary bonuses, I find no performance difference for subjects who receive the monetary gift of a 75% wage increase compared to a baseline of no gift. A comparable (with respect to the hourly wage) gift of leisure time, however, outperforms the monetary gift by an average of around 25% higher performance in the first period. The mechanism for this is an approximately 45% decrease in subjects' amount of on-the-job leisure (Internet) consumption during this time. Importantly, the total performance after both working periods is not substantially different between the money and the time gift treatment, despite the differences in the working time.

An online survey experiment among human resource managers complements the results with some external validity. Managers anticipate that a gift of leisure time at home will reduce on-the-job leisure consumption but they also state further advantages of a leisure time gift. Specifically, they expect an increase in efficiency, work commitment, health, satisfaction and felt appreciation compared to a gift of more money. Further control treatments varying the size of the leisure gift in the laboratory experiments, explore that granting leisure time consistently leads to a reduction of on-the-job leisure (Internet) instead of motivating subjects to work faster. However, the treatment effects do not vary with different sizes of the leisure time gift. Analyzing post-experimental questionnaire data, additional control treatments, and the online survey experiment, it seems that the above outlined theoretical idea of decreased marginal utility of leisure might not be the only one and that the role of reciprocity should not be neglected.

With these results, this study contributes to the literature on non-monetary bonuses, exploring a new bonus domain that is significantly different from other monetary and non-monetary bonus domains investigated by researchers in the past. The reduction of on-the-job leisure seems to be of practical importance, due to possibly high valuations of leisure time nowadays and possible distortionary behavior if employees are paid based solely on availability (see e.g. Chan 2018, Wiswall and Zafar 2018). Especially if employers have no information on how employees spend their working time, workers might have incentives to slack, reducing productivity.[8]

specific experimental design to reduce reciprocal intentions and focus on the substitution between leisure time at home and on the job.

[8] Another less investigated option to decrease on-the-job leisure, is to let employees go home after completing a project. This reduces duration of a working day (see, e.g., Englmaier et al. 2018).

The study may also contribute to the literature on inefficiently long working hours and productivity (see, e.g., Landers et al. 1996, Brachet et al. 2012, Pencavel 2015, Collewet and Sauermann 2017).[9] The consensus of these studies is that fatigue leads productivity to suffer under longer working hours. While fatigue is unlikely to influence the results of this study, results of this study can also be seen in light of this research strand. I, however, propose a different theoretical explanation to explain why a reduction in working hours can have positive effects on productivity.

4.2 Experimental Design and Implementation

The real-effort task of this laboratory experiment is a variation of the slider task by Gill and Prowse (2012). Four sliders are shown on the computer screen and have to be moved to a randomly predetermined position using the computer mouse. After all sliders are at the required position, the subject can proceed with the next four sliders and a different required position. This counts as one completed task and is referred to in the following as the subjects' performance. A further component is that subjects are able to browse the Internet during the task. At the beginning of every new task they can click the "timeout" button, which opens the Internet browser. Subjects can choose freely how to spend their time on the Internet and can minimize the Internet browser at any time to continue working.[10] Real-leisure alternatives in real-effort laboratory experiments are an important tool to study incentive effects, since subjects would otherwise work because they lack desirable alternatives (Corgnet et al. 2015, Araujo et al. 2016, Goerg et al. 2019). Internet consumption is a prominent real-leisure alternative, and furthermore it maximizes individual utility due to its fungibility.[11]

In the first part of the experiment, subjects become familiar with the task, which serves as a proxy for subjects' ability. Additionally, they learn how to access the Internet. In the second part, all subjects receive a notification that the experiment will last 30 minutes less (90 minutes in total) than stated in the invitations. Subjects then have to state their willingness-to-accept (henceforth called WTA) to stay 30 minutes longer in the lab.[12]

[9] Inefficiently long working hours have also received some public attention in the past with, for instance, Sweden starting a small pilot project implementing a six-hour instead of an eight-hour working day (*Washington Post*, April 21, 2017).
[10] See the appendix for the instructions (4.9.3), a screenshot of the working stage (4.9.4), and a screenshot of the Internet stage (4.9.5).
[11] The post-experimental questionnaire provides some evidence that subjects perceive the task as labor and that they use the Internet as leisure activity. Subjects evaluate different statements on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree) after the experiment. Across all treatments, they state low enjoyment of the task (mean = 2.44, SD = 1.76) and that they did not do anything productive during their internet time (mean =3.37, SD = 1.83). Both means are significantly different from the neutral response of 4 (Wilcoxon Signed-Rank Test, $p < 0.001$).
[12] To make this incentive compatible, I use the Becker-DeGroot-Marschak mechanism (Becker et al. 1964). One subject was randomly selected after each session to stay in the lab. Precise explanations, examples, control questions and the explicit

The third and main part of this experiment consists of two working periods. The task in each working period is identical to the task in the trial period, with subjects having the option of browsing the Internet. Initially, both working periods last 30 minutes. Subjects receive an unconditional, base compensation of €4 for both working periods together in addition to their show-up fee of €4. A possible unconditional additional compensation ("principal-first gift", Carpenter 2017) varies across treatments.[13] The treatments are as follows:

Baseline (N=43)

Subjects in the *Baseline* treatment first receive general information that the remainder of the experiment consists of two working periods of 30 minutes each. They are then notified that they will be compensated €4 for both working periods (base compensation), which together last a total of 60 minutes.

MoneyGift (N=42)

The *MoneyGift* treatment is the same as the *Baseline*, with the only difference being that subjects receive information that they will receive additional compensation of €6 for the first working period. In contrast to previous studies, the framing is neutral without any appreciation of the work subjects will do in the following period. It is also common knowledge that there will be no additional compensation in the second working period.

TimeGift25 (N=40)

In the *TimeGift25* treatment, subjects receive the same information and framing as in *MoneyGift*. The only difference is that instead of receiving an additional €6, the additional compensation for the first working period is that subjects can leave the laboratory 25 minutes earlier during the second working period.

I parameterize the experiment in a way that the gift treatments generate approximately the same hourly wage. In *MoneyGift* subjects receive €14 for 60 minutes of work and in *TimeGift25* subjects receive €8 for 35 minutes of work (= €13.7 for 60 minutes). Subjects were separated into two different rooms according to their treatment, and they were not aware of the

statement that it is not optimal to deviate from the true minimum value increases understanding of the mechanism. Instructions can be found in the Appendix 4.9.3.

[13] I deliberately choose the wording of "additional compensation" instead of "gift" or "reward" in the instructions to be as value-free as possible and not induce reciprocity intentions. Technically, however, the additional compensation can be seen as a gift or a reward. If anything, the effects of this study would be larger with a different wording. I further arguably reduce reciprocity intentions by not implementing a beneficiary from working.

other treatments. Thus, subjects in *MoneyGift* did not notice that subjects in *TimeGift25* were able to leave earlier. Subjects were also not allowed to do anything not associated with the experiment (e.g., read books or use smartphones).[14]

The experiment took place in July 2016 (*MoneyGift* and *TimeGift25*) and March 2017 (*Baseline*) with a total of 125 students (mean age = 23.69, 60.8% female, 46.4 % business or economics students), who were invited for a two-hour experiment using the online-platform ORSEE (Greiner 2015). Descriptive statistics about the experiment and the subject pool are shown in the Appendix Table 4.3A. The 24 sessions were evenly distributed among morning, noon and afternoon sessions and executed with zTree (Fischbacher 2007).

4.3 A Theoretical Framework

As a framework for organizing thoughts, this section develops a possible toy model for a hypothesis on the mechanism of this experiment. Using standard individual utility maximization, I derive a possible hypothesis for the experiment. Consider an agent with a quasi-linear utility function:

$$U = u_L(H + J) + m - c(J)$$

with $u_L' > 0, u_L'' < 0, c' > 0, c'' > 0, J \in [0,1]$. u_L denotes utility from leisure. Leisure can be consumed at home *(H)* and on-the-job *(J).* The regular working time is normalized to 1 such that the agent's actual working time is *(1-J).* The second part of the equation represents utility from money m and costs of consuming on-the-job leisure *c(J)* (e.g., potential sanctions when getting caught or deviations from work norms).[15]

Utility maximization with respect to on-the-job leisure *J* yields the following hypothesis:

Substitution Hypothesis: Additional leisure time at home reduces the marginal utility of on-the-job leisure and thus its consumption.

[14] Roughly every 10 minutes the experimenter entered the room to go through the laboratory and control for this.

[15] Remember that the experiment was designed to minimize possible effect from gift-exchange and a resulting reciprocal behavior by agents. However, for completeness, I provide a variant of a toy model using a reciprocity-based utility function in Appendix 4.9.1.

Proof:

The first order condition yields $\frac{\partial U}{\partial J} = u_L'(H + J) - c'(J)$.

It is now of interest how the optimal level of on-the-job leisure J^* changes with changes in leisure time at home H. By the implicit function theorem, the derivative with respect to leisure time at home is negative:

■

Due to the concave utility of leisure, the derivative of the optimal on-the-job leisure level with respect to leisure at home (*H*) is always negative. Additional leisure time at home, thus, decreases the marginal utility of on-the-job leisure, where the agent still faces costs of not working.[16] Therefore, on-the-job leisure consumption decreases, and work performance increases.

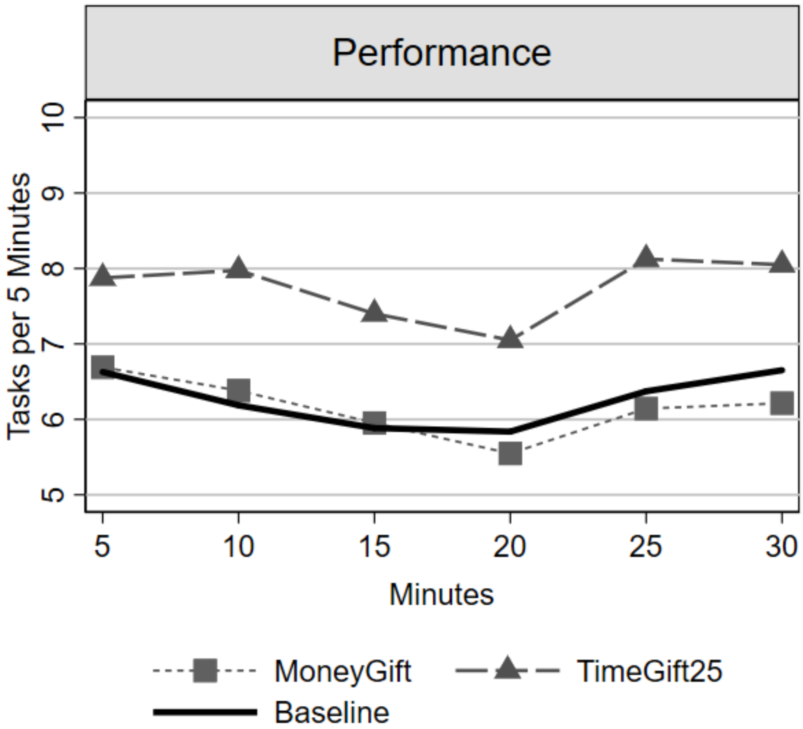
4.4 Results

4.4.1 Main Result

Figure 1 shows the distribution of completed tasks per 5 minutes in the first working period, and Table 4.4A in the Appendix displays descriptive statistics. Two findings become apparent. On the one hand, the monetary gift (*MoneyGift*) does not influence performance compared to the *Baseline* treatment.[17] Subjects in the *MoneyGift* treatment complete, on average, 36.93 tasks in the first working period, which is comparable to the *Baseline* of 37.56 tasks (Mann-Whitney-U (MWU) test, $p = 0.8846$).[18] On the other hand, the gift of more leisure time at home (*TimeGift25*) leads to an average of 46.48 completed tasks in the first period and, thus, outperforms *MoneyGift* and *Baseline* during the whole first working period (MWU test, $p = 0.0324$ and $p = 0.0754$, respectively).

[16] In the described experiment the costs are, for instance, the experimenter entering the room and discovering someone browsing the Internet or a deviation from a possible norm to work on a given task in an experiment. Arguably, the costs for not working are greater in organizations, which, however, would only increase the substitution effect.
[17] The high effort provision under the *Baseline* might be driven by some obligation to work (work norm) when subjects come to the laboratory.
[18] If not stated differently, all statistical tests in this study are two-sided.

Figure 4.1 – Completed Tasks per 5 Minutes in the First Working Period



Note: The Figure displays the average completed tasks per five-minute interval in the first working period.

The OLS regressions in Table 4.1 support these findings. In *TimeGift25*, subjects complete, on average, 8.9 tasks more than in the *Baseline* (p = 0.043, column 1) and 9.5 more tasks than in the *MoneyGift treatment* (Wald test, p = 0.0318, column 1). This represents a performance increase of approximately 25%. Controlling for socio-economic characteristics, different times of a day, an ability proxy from the trial period, and the individual valuation of time (WTA) does not alter these results (column 2).[19] Moreover, the cumulative distribution functions displayed in the Appendix Figure 4.7A show that the results are not driven by single subjects. Rather, the distribution in *TimeGift25* is shifted to the right in the direction of more tasks completed.

[19] Table 4.5A in the Appendix replicates Table 4.1 using log values. Moreover, because *MoneyGift* and *TimeGift25* were randomized within each session, I can run a regression with only these two treatments and control for session effects and also cluster on the session level. Again, this does not alter the results between these two treatments.

Table 4.1 – Main Treatment Effect – First Working Period

	(1) Total Tasks	(2) Total Tasks	(3) Total Internet (in sec.)	(4) Total Internet (in sec.)	(5) Total Tasks
MoneyGift	-0.630	1.217	62.838	30.664	2.203
	(4.316)	(4.282)	(125.392)	(123.159)	(1.640)
TimeGift25	8.917**	9.205**	-264.151**	-235.024*	1.650
	(4.370)	(4.279)	(126.968)	(123.071)	(1.665)
Internet Consumption					-0.032***
					(0.001)
Constant	37.558***	30.087***	488.934***	305.317	39.902***
	(3.034)	(12.972)	(88.143)	(373.074)	(4.982)
Controls	No	Yes	No	Yes	Yes
N of subjects	125	125	125	125	125
Observations	125	125	125	125	125
Adj. R^2	0.030	0.107	0.057	0.135	0.869

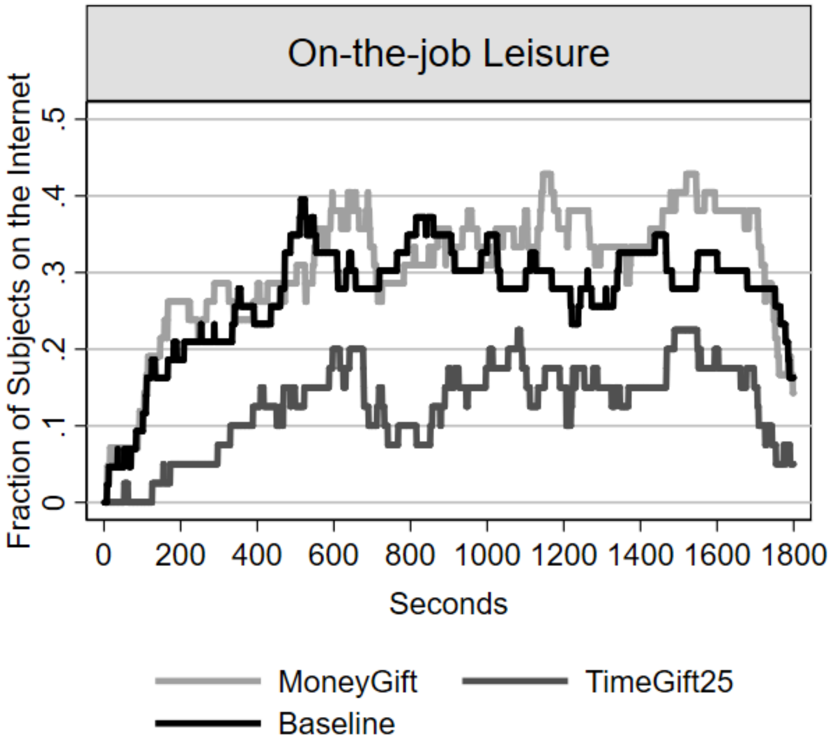
Note: The table reports coefficients from an OLS regression. The dependent variable is the number of completed tasks in the first working period in columns 1, 2, and 5. In columns 3 and 4, the dependent variable is seconds spent on the Internet in the first working period. Standard errors are in parentheses and clustered at the individual level. All results are compared to the *Baseline* treatment. *Internet Consumption* is the total amount of time (in seconds) the subjects spent on the Internet in the first working period. *Controls* include age, a gender dummy, a dummy for economics/business students, ability, dummies for the different times of the day, and the WTA. $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

4.4.2 Mechanism

There are two possible differences in behavior that could lead to these differences in performance. Either subjects work faster in *TimeGift25*, or they consume less on-the-job leisure (Internet). Concerning the former, subjects in *MoneyGift* need on average 35.63 seconds to complete one task. This is similar to the 35.55 seconds subjects under *TimeGift25* need to complete one task (MWU test, $p = 0.8280$).

Consumption of on-the-job leisure (Internet), however, is different across treatments and results in the differences in performance (Figure 4.2). Again, *TimeGift25* is clearly discernable from the other treatments in terms of the fraction of subjects per treatment who browse the Internet at each point in time during the 1,800 seconds in the first working period. Column 3 and 4 of Table 4.1 show the same regression as above with the average seconds spent using the Internet as the dependent variable. In the *TimeGift25* treatment, subjects browse the Internet for an average of 264.15 seconds fewer than in the *Baseline* (column 3). This represents a reduction of approximately 45%. Furthermore, it is significantly different from *MoneyGift* (Wald test, $p = 0.0117$, column 3). Controlling for the total amount of seconds on the Internet in an OLS regression, with the amount of completed tasks as dependent variable (column 5), results in insignificant and comparable treatment effects. Hence, the treatment effect is nearly entirely driven by a reduction of on-the-job leisure consumption. This in in line with the possible theoretical explanation of decreased marginal utility of leisure time and the resulting substitution effect.

Figure 4.2 – Fraction of Subjects on the Internet per Second in the First Working Period



Note: The Figure displays the fraction of subjects on the Internet depending on the time (in seconds) in the first working period.

4.4.3 Economic Efficiency

As subjects with the *TimeGift25* have less working time, it is an important question whether this pays off for an employer. In *TimeGift25* subjects only work 35 minutes in total. They nevertheless complete an average of 53.35 tasks after both working periods, which is not substantially different from the average of 64.62 completed tasks in *MoneyGift*, where subjects have 60 minutes of working time (MWU test, $p = 0.1557$) and higher payments. Subjects thus have 41.67% less working time but an only 17.41% lower total output. The average number of completed tasks per working minute further reflects this. On average, subjects complete 1.52 tasks per work minute in *TimeGift25* and 1.08 tasks per work minute in *MoneyGift* (MWU test, $p = 0.0039$). Importantly, subjects in *TimeGift25* are also paid €6 less. This leads to an average cost of €0.22 per completed task in *MoneyGift* and only €0.15 in *TimeGift25*. Under the described circumstances and conditional on giving employees' an unexpected gift, it would have been efficient to provide the time gift up until the point where the task itself is worth below €0.53 per unit of output.[20] Hence, from an employer's perspective it can be efficient to consider time as a potential bonus domain instead of money, even though this entails allowing the workforce to leave earlier.

Given the total output (1st and 2nd Period, Table 4.4A) the *Baseline* seems also to be an alternative to the gift treatments. However, this depends crucially on the parametrization of the main treatments of this experiment.[21] Therefore, I conduct the treatment *TimeGift10* (N = 39) in which everything remains similar to *TimeGift25* except that subject now can leave the laboratory only 10 minutes earlier. The gift of 10 minutes more leisure (and thus only 50 minutes of working time) results in an average of 79.41 solved tasks in both working periods and therefore even slightly more than in the *Baseline* of 60 minutes working time with an average of 77.30 tasks (MWU test, $p = 0.5491$).[22] With this parametrization and given the presented environment it would thus be optimal for both sides to let subjects leave earlier.

4.5 Survey Experiment with Human Resource Managers

To complement the experimental analysis and collect some evidence on the external validity, I ran an online survey experiment in cooperation with the German Association for Human Resource Management (DGFP) in July 2019. Since a big part of a human resource (HR) manager's job is to increase the performance of employees by designing effective bonus and compensation schemes, this is an optimal group to survey about the efficacy of time versus money bonus payments (gifts).[23] In total 141 managers participated in the survey (overall mean

[20] With *TimeGift25*, the "employer" saves €6 but also receives on average 11.26 fewer tasks. €6/11.26 tasks represents the upper bound, up to which it is still efficient to provide a gift.

[21] Moreover, this interpretation leaves out any discussion on possible effects on employees' happiness, satisfaction, health, work commitment, felt appreciation and the reduction of other economic costs such as rents and electricity for longer working hours. The high amount of baseline work performance might also be an experimental artifact due to possible strong work norms in the laboratory. From the perspective of a social planner, who also values the gained freedom of the employees, the implemented parameterization is already efficient. Employees' value the 25 minutes leaving earlier with €5.97 (calculated from the elicited WTA) and thus 75% of their initial baseline wage. The employer pays approx. €0.10 per correct task in the Baseline and the difference in output between *Baseline* and *TimeGift25* is overall 23.91 tasks. Hence, the employer values the difference with €2.39 and thus less than the additional value of the employees for leaving earlier.

[22] Again using the elicited WTAs, subjects gain a utility of €2.39 from leaving the laboratory 10 minutes earlier.

[23] The approach is similar to Heinz et al. (2017). DellaVigna and Pope (2018) show that experts do a good job in forecasting certain outcomes concerning incentives.

age = 44.33).[24] The Appendix 4.9.6 displays the precise survey items, as well as all means and standard deviations.

At the beginning of the survey, managers had to evaluate the novelty and value of a time gift and a money gift. [25] The managers generally pronounce what has already been stated in the introduction – a time gift is a rather novel form of a bonus payment compared to a money gift (Wilcoxon Signed-Rank (WSR) Test, $p < 0.001$). Moreover, they evaluate the time gift as a better form of bonus payment than a money gift (WSR, $p < 0.001$).

I then randomly assign 68 managers to a survey with a focus on time gifts[26] and 73 managers to a survey with a focus on money gifts[27]. Managers had to evaluate different statements (on a scale from 1 = totally agree to 7 = totally disagree) starting with "*For the rewarded employees/colleagues …*".

Concerning the different effects on employees granted with one of the two gift forms, they expect significant differences (Figure 4.3). They anticipate a greater positive influence on an employees' health, work commitment and work satisfaction when they are granted with a time gift rather than with a money gift (MWU, all $p < 0.001$). They also expect a higher felt appreciation from the employer when employees receive a gift of time, which may indicate possible reciprocal behavior of employees (MWU, $p < 0.001$). Importantly, the HR managers also anticipate the experimental outcome of a greater efficiency in case employees receive a time gift and fewer on-the-job leisure consumption compared to a money gift (MWU, all $p < 0.001$). They further do not expect greater stress among those employees (MWU, $p = 0.9513$), something which could be a drawback of a time gift.

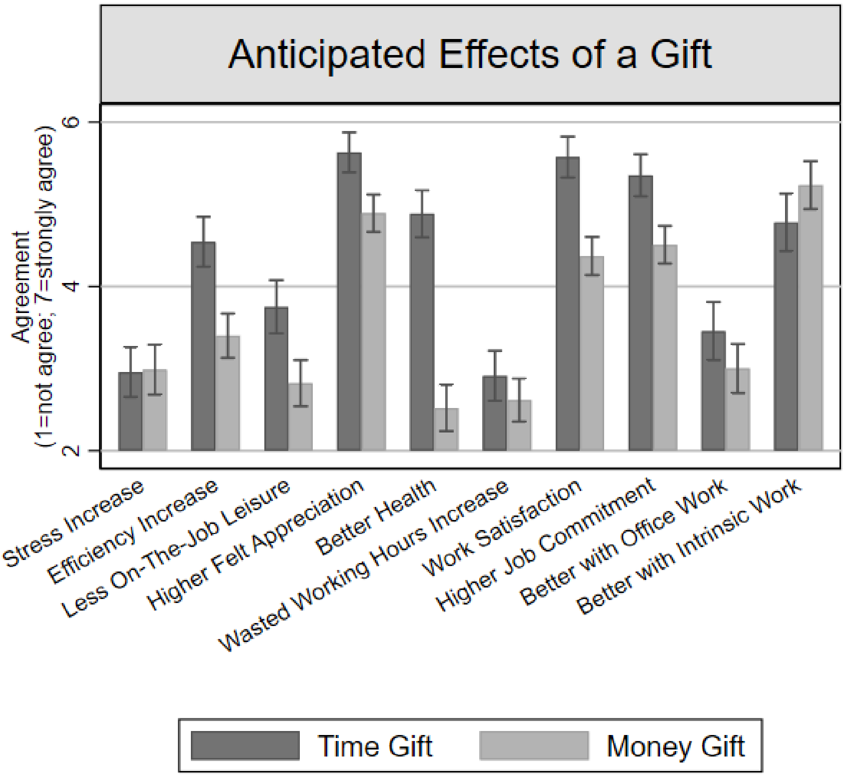
[24] The survey data represents a broad sample of companies. Out of 20 possible categories to which industry a company belongs, all categories were selected, with industrial goods and chemicals appearing the most frequent (approx. 10% each). The data further represents all sorts of company sizes. For example, 8.51% state that their company has below 100 employees and 34.75% state that their company has above 5000 employees.

[25] I define the time gift as a leisure bonus which is "*a leisure spot-bonus (-award,-reward), thus an immediate, onetime permission to work less without wage reductions"* and money gift as a money bonus which is *"a financial spot-bonus (-award,-reward), thus an immediate, onetime additional financial payment*".

[26] The precise framing was: "*What do you think would happen if your employees/colleagues were given a "Leisure-Bonus" (compared to a situation where there is no bonus at all)?*"

[27] The precise framing was: "*What do you think would happen if your employees/colleagues were given a "Money-Bonus" (compared to a situation where there is no bonus at all)?*"

Figure 4.3 – HR Managers' Anticipation of Different Outcomes



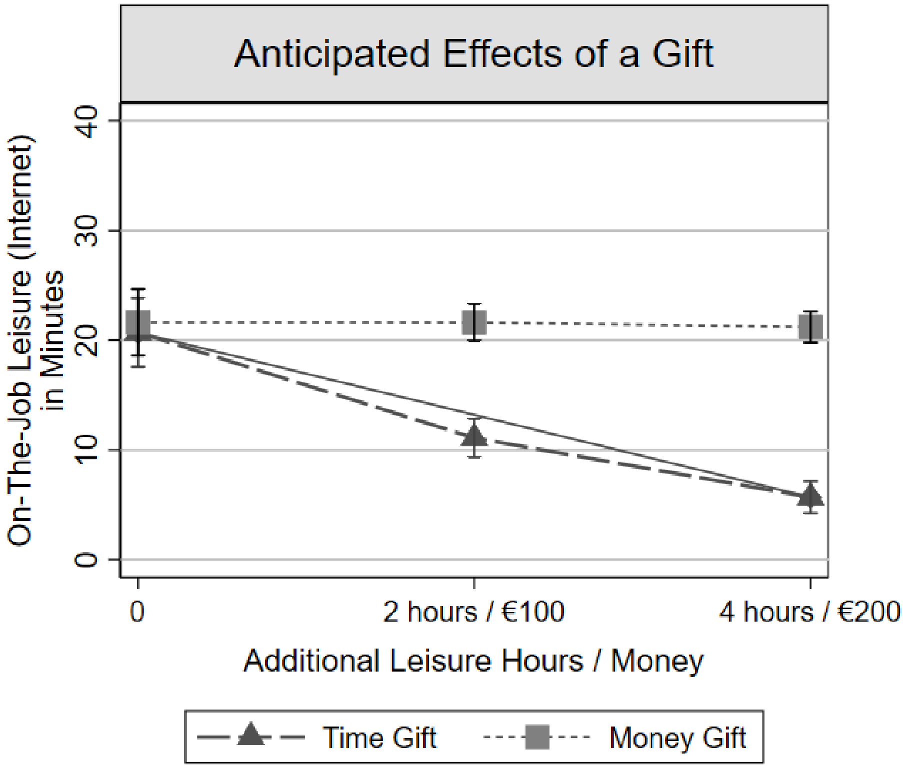
Note: The figure displays mean outcomes from an online survey experiment with HR managers (N=141). It shows the between comparison on the managers expectations about effects of either a time (N=68) or a money (N=73) gift on different outcomes for the rewarded employees/colleagues. 90% confidence bands are displayed.

Interestingly, they do not only anticipate the difference in efficiency, but also that it will result because of less on-the-job leisure. I ask the managers about their anticipated on-the-job leisure consumption (private Internet usage) for an average employee with 8 hours of office working time (mean = 21.18 minutes, SD = 15.64). Afterwards, I used the randomly assigned groups from the beginning and ask about the amount of on-the-job leisure consumption (private Internet usage) in case of a 2 and 4 hour gift of leisure time the following day (survey with time focus) or a €100 and €200 gift (survey with money focus).[28] Figure 4.4 shows the anticipation of managers. Managers answering the questions for the time gift anticipate the significant decrease in on-the-job leisure consumption (private Internet usage) similar to the experimental results (One-Way Repeated Measure ANOVA, $p < 0.001$). Importantly, they also anticipate the decrease to be sensitive to the amount of time given. The other group of managers anticipate

[28] One item for managers in the time gift group was for instance: "*An employee/colleague receives a one-time "Leisure-Bonus" of 2 hours, which he can leave earlier from work the next day. What do you think, how many minutes of the remaining 6 hours of work would this employee spend privately on the Internet?*"

that a money gift does not change on-the-job leisure consumption at all (One-Way Repeated Measure ANOVA, $p = 0.734$).

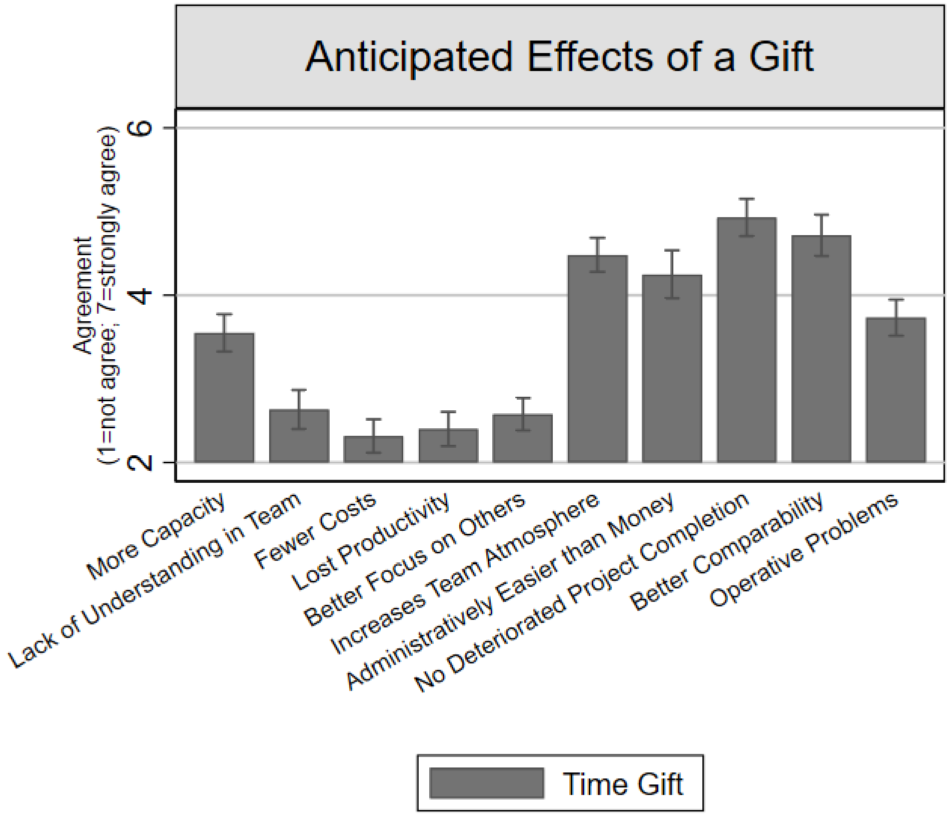
Figure 4.4 - HR Managers' Anticipation of Changes in On-The-Job Leisure



Note: The figure displays mean expectations from an online survey with HR-managers (N=141). Managers are asked in a between-subject design about their anticipation of how on-the-job leisure develops with different sizes of either a time or a money gift. 90% confidence bands are displayed. The solid line displays a linear function from 0 hours to 4 hours of a time gift.

I further ask questions (to managers of both surveys similarly) about the impact on the company environment when employees receive a gift of more leisure time (Figure 4.5). Here it seems reassuring that many of the possible practical concerns seem not to be vast concerns of the managers. Importantly, the HR managers even anticipate an increase in the team atmosphere (WSR against neutral response (4), $p < 0.001$) and that there would be no lack of understanding within the team (WSR against neutral response (4), $p < 0.001$). Moreover, they also do not think that fewer projects will be completed (WSR against neutral response (4), $p < 0.001$) or that the productivity of the employee rewarded with a time gift will suffer due to less working hours (WSR against neutral response (4), $p < 0.001$).

Figure 4.5 – HR Managers' Anticipation of Different Outcomes



Note: The figure displays mean outcomes from an online survey with HR-managers (N=141). It shows statements asked to all managers on the effect of a time gift on the work enviornment and administrative concerns. 90% confidence bands are displayed.

4.6 Discussion of Results

The results presented above show that subjects in the laboratory deliver significantly higher performance in *TimeGift25* through the mechanism of less on-the-job leisure. These results are in line with a possible theoretical explanation of decreased marginal utility of on-the-job leisure when granted with additional leisure time at home. HR managers anticipate this mechanism in an online survey experiment but also point out further influences of a time gift that may demonstrate additional possible explanations for the results.

Using ex-post questionnaire data from the experiment as well as additional control treatments in the laboratory, this section discusses the concavity assumption necessary for the idea of decreased marginal utility of leisure, the role of reciprocity, and further potential alternative explanations for the performance effect.

4.6.1 Concavity of Leisure Time

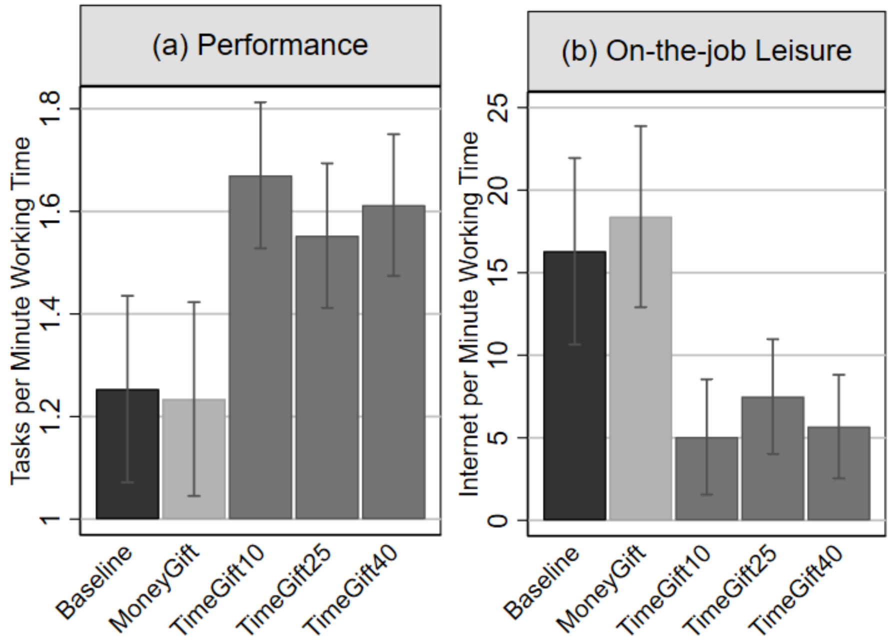
The theoretical framework in the beginning (section 4.3) helps to explain why a gift of more leisure time could lead to a decrease in on-the-job leisure (as opposed to working faster). However, for the resulting substitution between leisure time at home and on-the-job leisure a key assumption is a concave utility of overall leisure time.

This assumption seems to find evidence in the recent literature (Mas and Pallais 2019). To explore this further in the current setting, I run the treatment *TimeGift40* (N = 41) in which everything remains the same except that subjects now can leave the laboratory 40 minutes earlier. Figure 4.6 displays the treatments with varying size of the time gift (10 minutes = *TimeGift10*, 25 minutes = *TimeGift25*, 40 minutes = *TimeGift40*). It shows no notable difference in the first working period concerning the amount of completed tasks per working minute or the time subjects browse the Internet per working minute. Subjects complete on average 1.67 tasks per working minute in *TimeGift10*, 1.61 tasks in *TimeGift40*, and 1.55 tasks in *TimeGift25*.[29] The differences between the time gift treatments are not statistically different from each other (Kruskal-Wallis Test, $p = 0.5509$). Concerning the Internet time per working minute, the time gift treatments are also not significantly different from each other (Kruskal-Wallis Test, $p = 0.5147$).[30]

[29] Table 4.6A in the Appendix shows descriptive statistics about the subject pool of all control treatments. Table 4.7A in the Appendix shows descriptive statistics about the main outcomes of all control treatments.

[30] The finding that real-effort outcomes in the laboratory are not very sensitive to different sizes of incentives is not new to the literature. Araujo et al. (2016), for instance, find no difference in the outcomes when subjects face different performance pay incentives in the laboratory. Sliwka and Werner (2017) find differences in performance in a dynamic real-effort laboratory experiment when subjects receive different sizes of a monetary gift over eight rounds. Importantly, in their setting it is also not the size of the gift that drives the performance effect, but the fact that there is a change in the size of the gift for the subjects.

Figure 4.6 – Additional Control Treatments Varying the Size of the Time Gift



Note: The figure displays results per minute of working time in the first period for different treatments. Panel (a) shows the average amount of completed tasks. Panel (b) shows seconds in the internet. 90% confidence bands are displayed.

However, in all three time gift treatments subjects react again by lowering their Internet consumption, rather than by working faster. All time gift treatments result in less on-the-job leisure per working minute compared to *MoneyGift* (MWU tests, all $p < 0.01$), while the time subjects need per task is not different (MWU tests, all $p > 0.1$).

To conclude, varying the size of the leisure time gift does not alter the experimental results. Thus, the additional treatments do not provide empirical support for the substitution of leisure at home and leisure on-the-job as a single possible theoretical explanation and demonstrate that further effects might be present. This is also in line with the survey experiment among HR managers. Although they do in fact anticipate a (slightly) concave decrease of on-the-job leisure with an increase in the size of the leisure time gift (Figure 4.4), they also anticipate possible further influences on different employee outcomes, which could be a result of other effects besides the substitution (Figure 4.3). Nevertheless, it is noteworthy that the mechanism for any performance improvements in a time gift treatment compared to the monetary gift treatment are always due to a reduction in on-the-job leisure rather than a higher intensity of work.

4.6.2 Reciprocity

Most of the studies on unconditional monetary and non-monetary rewards refer to the concept of reciprocity to explain performance effects. They argue that employees want to return the favor of these rewards.[31] The results from the survey experiment with HR managers provide some initial evidence about the possible role of reciprocity with a gift of more leisure time. Managers expect an increase in the felt appreciation of employees, which can be a driver of reciprocal behavior. They also expected an increase in work satisfaction, commitment to work, and general efficiency - components possibly driven by reciprocal reactions. Ex-ante, the effect of reciprocity in the laboratory study presented above might be considerably low (neutral framing in the experimental design, task without real impact, no beneficiary from work). However, given the performance effects of the control treatments described in section 6.1, which are independent of the size of the time gift, and the survey results, it is worthwhile to investigate the mechanism of reciprocity further.

I use an ex-post questionnaire of the gift treatments and an additional treatment to investigate this in more detail. Specifically, I run the control treatment *Baseline35* to disentangle the theoretical explanations (N = 42). This treatment is similar to the *Baseline*, except that subjects know right from the invitation that the experiment will be 25 minutes shorter.[32] They then have to work for 30 minutes in the first working period and 5 minutes in the second without framing it as an additional compensation and, thus, the treatment is intended to further reduce reciprocal behavior. In the same line, there is no deviation from an a priori reference point concerning the length of the experiment. This also reduces a possible surprise effect, which is often argued to result in reciprocal behavior (Rogers and Frey 2015, Khalmetski et al. 2015, Bradler et al. 2016, Macera and te Velde 2018).

Table 4.7A in the Appendix displays the key outcome measures of this additional control treatment. It appears that *Baseline35* is not clearly distinguishable from, *TimeGift25* nor *MoneyGift*. In fact, while subjects complete 44.74 tasks, which is 21.15% more than in *MoneyGift* (MWU test, $p = 0.0980$), they only complete 3.74% fewer tasks than in *TimeGift25* (MWU test, $p = 0.9883$). Concerning their on-the-job leisure consumption (Internet), however,

[31] Furthermore, studies on unconditional monetary bonuses with close to zero average treatment effects find heterogeneous treatment effects that vary with the subjects' reciprocity (see, e.g., Englmaier and Leider 2012, Carpenter 2016).

[32] To avoid selection issues and still reduce possible reciprocity due to a gift of leisure time, I sent an email to all participants one day before the experiment stating that the experiment will last approximately 25 minutes less. As showing up to the experiment (with 2 no shows among 46 subjects (2 of which were reserve candidates) in *Baseline35* and 5 no shows among 137 subjects (5 of which were reserve candidates) in the three main treatments) did not differ among the treatments (χ^2 (1), $p = 0.8308$), I can assume that this procedure did not create any further selection problems from the beginning. Note that they are (as in the other treatments) additionally notified in the experiment that it will last additional 30 minutes less because of the elicitation of the WTA.

subjects browse the Internet in *Baseline35* for 339.42 seconds which is 38.49% less than in *MoneyGift* (MWU test, $p = 0.1898$), but 51% more than in *TimeGift25* (MWU test, $p = 0.4978$).

As the evidence of the additional control treatment seems to be inconclusive, I further use post-experimental questionnaire data from a hypothetical investment game, which is a proxy for subjects' reciprocal inclinations (Falk et al. 2016).[33] All subjects are in the role of the second-mover and have to decide how much they would return to a person who sent (after the amount was tripled) €15, €30, €45 and €60. Table 4.2 displays regression results pooling all treatments with a time gift (*TimeGift10*, *TimeGift25*, *TimeGift40*) together (*TimeGiftAll*) and using the average completed task per working minute in the first working period as dependent variable.[34] Column 1 and 2 interact the gift treatments with the standardized average amount returned back and column 3 and 4 interact the gift treatments with a dummy variable indicating whether the average amount returned back is above the median (€18.125).

I do not find that reciprocal inclinations influence performance in the *Baseline.* Point estimates are negative and when investigating the median split of reciprocal inclinations, the effect is even marginally significantly negative. Point estimates for the interaction of the reciprocity proxy with *MoneyGift* and *TimeGiftAll* are positive in all specifications with the estimates of *TimeGiftAll* always being greater than that of *MoneyGift*. Moreover, depending on the specification, the interactions are significantly different from 0 for *TimeGiftAll*. For *TimeGiftAll* the interaction effect is, for instance in column 1, approximately 68.6% the size of the treatment effect. Thus, the treatment effect is more than halved for subjects with one standard deviation below the mean amount returned back (mean responsiveness to gifts/ reciprocal inclinations). In column 4, the treatment effect of *TimeGiftAll* is approximately 2.5 times larger for subjects with an above median level of reciprocal inclinations. Moreover, subjects with an above median level of reciprocal inclinations increase performance approximately 2.5 times more with a gift of time than with a gift of money. Hence, I do find some evidence that reciprocal subjects behave differently and that they tend to reciprocate a time gift to a stronger extent than a money gift.

[33] As this is non-incentivized data it might be viewed with more caution than incentivized data. However, it is not clear that subjects have incentives to deviate from their true preferences although questions are only hypothetical. Sliwka and Werner (2017) use a similar analysis (although a different non-incentivized reciprocity question) to show that their effects are driven by reciprocal subjects.

[34] Table 4.8A in the Appendix replicates Table 4.2 but shows effects for all treatments separately.

Table 4.2 – The Role of Reciprocity

	(1) Tasks per Minute of Working Time	(2) Tasks per Minute of Working Time	(3) Tasks per Minute of Working Time	(4) Tasks per Minute of Working Time
MoneyGift	-0.017	-0.006	-0.098	-0.078
	(0.154)	(0.151)	(0.1998	(0.191)
TimeGiftAll	0.361***	0.381***	0.126	0.172
	(0.116)	(0.123)	(0.155)	(0.155)
Reciprocity	-0.173	-0.166		
	(0.135)	(0.131)		
MoneyGift x Reciprocity	0.172	0.165		
	(0.175)	(0.169)		
TimeGiftAll x Reciprocity	0.247*	0.224		
	(0.144)	(0.143)		
Above Median Reciprocity			-0.393*	-0.374*
			(0.210)	(0.206)
MoneyGift x Above Median Reciprocity			0.191	0.176
			(0.305)	(0.295)
TimeGiftAll x Above Median Reciprocity			0.486**	0.443*
			(0.232)	(0.230)
Constant	1.247*	1.563***	1.435***	1.707***
	(0.144)	(0.350)	(0.133)	(0.350)
Controls	No	Yes	No	Yes
N of subjects	205	205	205	205
Observations	205	205	205	205
R^2	0.104	0.132	0.110	0.139

Note: The table reports coefficients from an OLS regression. The dependent variable is the number of completed tasks per working minute in the first working period. *TimeGiftAll* pools all treatments with a time gift (TimeGift10, TimeGift25, TimeGift40). Columns 1 and 2 represent different standardized interactions with the treatment variables. *Reciprocity* is the standardized average investment subjects state that they would return in a hypothetical post-experimental Investment Game. *Above Median Reciprocity* is a dummy variable indicating whether the subjects' reciprocity (measured by the Investment Game) is above or below the median of the subject pool studied in this table. Standard errors are in parentheses and clustered at the individual level. Everything is compared to the *Baseline* treatment. *Controls* include age, a gender dummy, a dummy for economics/business students, ability, dummies for the different times of a day, and the WTA. $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

4.6.3 Possibility to Work Faster

The above clearly shows that subjects perform better with a gift of more leisure time at home and that the mechanism for this is always a reduction of their on-the-job leisure (Internet) consumption during the working time. It is now important to show, that subjects could in principle also react on a different margin, e.g. working faster. Therefore, I conduct the *MoneyPfP (N = 39)* treatment, in which everything was held constant as in the *Baseline*, but subjects receive €0.10 per completed task in the first working period. This treatment, with an arguable higher incentive, shows that it is actually possible to work more quickly, demonstrating that a reduction in Internet consumption is not the only possible mechanism to complete more tasks. Subjects with performance pay need, on average, 31.38 seconds per task. This is significantly less time compared to the *TimeGift25* treatment, with 35.55 seconds per task (MWU test, $p = 0.0424$) and also compared to the other two treatments with a time gift (MWU test, $p = 0.0292$ for *TimGift40*, $p = 0.0713$ for *TimeGift10*).

4.6.4 Fatigue

A further explanation for the increased performance with time gifts, in addition to substitution between leisure time at home and on-the-job as well as reciprocity, could be fatigue. Studies on the optimal length of a workday argue that longer working hours reduce productivity due to fatigue effects (see, e.g., Brachet et al. 2012). They find that employees get tired from long working days and are then less productive. It might thus be the case that subjects in the experiment rationally expect to exhaust their resources and get tired after the 60-minutes working period. Subjects in *TimeGift25* then might shift their resources from the additional leisure time they just received into the first working period and work harder.[35] To investigate this, I use the performance from the *MoneyPfP* treatment. On average, subjects complete 58.36 tasks, which is 25.56% more tasks in the first working period than under *TimeGift25* (MWU test, $p = 0.0014$). Table 4.7A shows the remaining descriptive statistics compared to the other treatments. This is contrary to the argument that being able to leave earlier gives subjects in *TimeGift25* more energy to exert during their working time as subjects in *MoneyPfP*, who have to stay longer than in *TimeGift25*, also work harder.

[35] It might also be that subjects expect to get bored after 60 minutes of working and, thus, slow down in the first working period.

4.6.5 Preference and Valuation of Leisure

Another potential explanation could be that leisure time is more preferred or valued than money and thus subjects increase performance. Yet after the second working period, 75% of the subjects in *TimeGift25* state that they actually would have preferred a monetary compensation for this experiment, whereas only 21.4% of the subjects in *MoneyGift* state that they would have preferred time compensation ($\chi^2(1)$, $p < 0.001$).[36]

Moreover, controlling for the incentive-compatible elicited WTA (as a proxy for the opportunity costs of time) does not alter the results in Table 4.1.

4.7 Conclusion

Research on employees' motivation and performance is an essential part of business and economics. This study investigates a novel way to incentivize employees: additional leisure time at home. This domain is practically interesting as it complements the current rise of alternative work arrangements, and potentially addresses the problem that employees frequently engage in non-productive tasks or go on the Internet during work hours. From an academic perspective, this paper is of particular interest for two reasons. First, because of certain unique characteristics of the time domain (for instance fungibility and divisibility) and, second, for the mechanism through which a bonus of more leisure time works – a reduction in on-the-job leisure as distortionary behavior influencing the intensive margin of labor supply.

In the laboratory experiment of this study, an unconditional gift of leisure time reduces on-the-job leisure consumption and increases performance compared to an unconditional monetary gift. The mechanism of a reduction of on-the-job leisure instead of working faster is the same across all treatments where subjects receive different sizes of leisure time at home as a gift. To provide some external validity, an online survey experiment among human resource managers reveals that they anticipate the decrease of on-the-job leisure when granting a gift of more leisure time to employees compared to a monetary gift. Importantly, the survey also shows further advantages and impacts of time gifts. Managers, for instance, anticipate an increase in efficiency, work satisfaction, commitment to work and felt appreciation when employees receive a gift of more leisure time.

Post-experimental questionnaire data, additional control treatments in the laboratory and the results from the online survey experiment demonstrate that the initial possible theoretical

[36] This is a common finding using non-monetary domains for rewards (see, e.g. Kube et al. 2012).

explanation of decreased marginal utility of on-the-job leisure when receiving more leisure time at home might not be the only explanation for the result. In fact, the role of reciprocity should not be neglected.

The performance effect of leisure time bonuses, however, might be dependent on the different type of work tasks not investigated in this study. Heterogeneous effects concerning the intrinsic value of the task might be possible and also stated by HR managers in the survey. A bonus of leisure time while working on an intrinsically motivating task could be seen as negative and demotivating. Similar effects could occur depending on the current stress level of the employee, if, for example, leaving earlier would result in more work the following working day. Additionally, it might not be optimal to reduce on-the-job leisure to zero because of positive spillover effects, for instance, on concentration during the work day (Mednick et al. 2002), increased earnings due to greater networking activities (Hamermesh 1990), or increased knowledge flows between co-workers (Sandvik et al. 2019).

This study makes use of the advantages of laboratory experiments that hold as many factors as possible constant and generate precise data to explore the root of employees' behavior. Within these constraints, and with the above-mentioned limitations in mind, leisure time as a gift increases performance by reducing on-the-job leisure (a distortionary behavior) and is not strictly dominated by the monetary gift in terms of the eventual output. This has important implications for the design of organizations and bonus schemes. Taking into account what could be done with bonuses of leisure time under strategic workforce planning and the possible spillover benefits, for instance on health-related issues (Bannai and Tamakoshi 2014) and happiness (Whillans et al. 2017), it is a noteworthy alternative to monetary bonuses and the various forms of non-monetary bonuses that are currently used in practice.

4.8 References of Chapter 4

Akerlof, George. “Labor Contracts as Partial Gift Exchange.” *The Quarterly Journal of Economics* 97.4 (1982): 543-569.

Araujo, Felipe A., Erin Carbone, Lynn Conell-Price, Marli W. Dunietz, Ania Jaroszewicz, Rachel Landsman, Diego Lamé, Lise Vesterlund, Stephanie W. Wang, and Alistair J. Wilson. "The slider task: an example of restricted inference on incentive effects." *Journal of the Economic Science Association* 2.1 (2016): 1-12.

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. “Awards Unbundled: Evidence from a Natural Field Experiment.” *Journal of Economic Behavior & Organizations* 100 (2014): 44-63.

Bannai, Akira, and Akiko Tamakoshi. "The association between long working hours and health: a systematic review of epidemiological evidence." *Scandinavian Journal of Work, Environment & Health* 40.1 (2014): 5-18.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak. "Measuring utility by a single-response sequential method." *Behavioral Science* 9.3 (1964): 226-232.

Bloom, Nicholas, James Liang, John Roberts, Zhichun Jenny Ying. "Does working from home work? Evidence from a Chinese experiment." *The Quarterly Journal of Economics* 130.1 (2015): 165-218.

Bonner, Sarah E., and Geoffrey B. Sprinkle. “The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research.” *Accounting, Organizations and Society* 27.4-5 (2002): 303-345.

Brachet, Tanguy, Guy David and Andrea M. Drechsler. “The Effect of Shift Structure on Performance.” *American Economic Journal: Applied Economics* 4.2 (2012): 219-246.

Bradler, Christiane, and Susanne Neckermann. “The Magic of the Personal Touch: Field Experimental Evidence on Money and Appreciation as Gifts.” *The Scandinavian Journal of Economics* 121.3 (2019): 1189-1221.

Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non. "Employee recognition and performance: A field experiment." *Management Science* 62.11 (2016): 3085-3099.

Brown, Alexander L., Jonathan Meer and J. Forrest Williams. “Why Do People Volunteer? An Experimental Analysis of Preferences for Time Donations.” *Management Science*. 65.4 (2019): 1455-1947.

Carpenter, Jeffrey. "The labor supply of fixed-wage workers: estimates from a real effort experiment." *European Economic Review* 89 (2016): 85-95.

Carpenter, Jeffrey. “The sequencing of gift exchange: A field trial.” *Journal of Economic Behavior & Organization* 139 (2017): 26-31.

Chan, David C. “The Efficiency of Slacking: Evidence From The Emergency Department.” *Econometrica* 86.3 (2018): 997-1030.

Charness, Gary. "Attribution and reciprocity in an experimental labor market." *Journal of Labor Economics* 22.3 (2004): 665-688.

Charness, Gary, David Masclet, Marie Claire Villeval. "The Dark Side of Competition for Status." *Management Science* 60.1 (2014): 38-55.

Cohn, Alain, Ernst Fehr, and Lorenz Goette. "Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment." *Management Science* 61.8 (2014): 1777-1794.

Collewet, Marion and Jan Sauermann. "Working hours and productivity." *Labour Economics* 47 (2017): 96-106.

Conner, Ceryl. "Wasting Time At Work: The Epidemic Continuous." (2015) Retrieved from https://www.forbes.com/sites/cherylsnappconner/2015/07/31/wasting-time-at-work-the-epidemic-continues/#47bdbbe91d94

Corgnet, Brice, Roberto Hernán-González, and Eric Schniter. "Why real leisure really matters: Incentive effects on real effort in the laboratory." *Experimental Economics* 18.2 (2015): 284-301.

D'Abate, Caroline P., and Erik R. Eddy. "Engaging in personal business on the job: Extending the presenteeism construct." *Human Resource Development Quarterly* 18.3 (2007): 361-383.

Danilov, Anastasia, and Timo Vogelsang. "Time for helping." *Journal of the Economic Science Association* 2.1 (2016): 36-47.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, Gautam Rao. "Estimating Social Preferences and Gift Exchange at Work." *NBER Working Paper* No. 22043 (2016).

DellaVigna, Stefano and Devin Pope. "What Motivates Effort? Evidence and Expert Forecasts." *Review of Economic Studies* 85.2 (2018): 1029-1069.

Englmaier, Florian, and Stephen Leider. "Managerial payoff and gift exchange in the field." *CESifo Working Paper no.3707* (2012).

Englmaier, Florian, Stefan Grimm, David Schindler, and Simeon Schudy. "The Effect of Incentives in Non-Routine Analytical Team Tasks – Evidence from a Field Experiment." *CESifo Working Paper* No. 6903 (2018).

Esteves-Sorenson, Constança. "Gift Exchange in the Workplace: Addressing the Conflicting Evidence with a Careful Test." *Management Science* 64.9 (2018): 4365-4388.

Falk, Armin. "Gift Exchange in the Field." *Econometrica* 75.5 (2007): 1501-1511.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. „The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *IZA Working Paper No.9674* (2016).

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. "Does fairness prevent market clearing? An experimental investigation." *The Quarterly Journal of Economics* 108.2 (1993): 437-459.

Festjens, Anouk, Sabrina Bruyneel, Enrico Diecidue, and Siegfried Dewitte. "Time-based versus money-based decision making under risk: An experimental investigation." *Journal of Economic Psychology* 50 (2015): 52-72.

Fischbacher, Urs. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10.2 (2007): 171-178.

Gallus, Jana. "Fostering Public Good Contribution with Symbolic Awards: A Large-Scale Natural Field Experiment at Wikipedia." *Management Science* 63.12 (2017): 3999-4015.

Gibbs, Michael. “Past, Present and Future Compensation Research: Economist Perspectives.” *Compensation & Benefits Review* (2017):1-10.

Gilchrist, Duncan S., Michael Luca and Deepak Malhotra. “When 3+1>4: Gift Structure and Reciprocity in the Field.” *Management Science* 62.9 (2016): 2639-2650.

Gill, David, and Victoria Prowse. "A structural analysis of disappointment aversion in a real effort competition." *The American Economic Review* 102.1 (2012): 469-503.

Gneezy, Uri, and John A. List. “Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments.” *Econometrica* 74.5 (2006): 1365-1384.

Goerg, Sebastian J., Sebastian Kube, Jonas Radbruch. “The Effectiveness of Incentive Schemes in the Presence of Implicit Effort Costs.” *Management Science* (2019).

Greiner, Ben. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association* 1.1 (2015): 114-125.

Hamermesh, Daniel S. “Shirking or Productivity Schmoozing: Wages and the Allocation of Time at Work.” *ILR Review* (1990) 43.3: 121-133.

Hannan, R. Lynn, John H. Kagel, and Donald V. Moser. "Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences, and effort requests on behavior." *Journal of Labor Economics* 20.4 (2002): 923-951.

Heckman, James J. “What Has Been Learned About Labor Supply in the Past Twenty Year?” *American Economic Review P&P* 83.2 (1993):116-121.

Heinz, Matthias, Sabrina Jaworrek, Vanessa Mertins, Heiner Schumacher and Matthias Sutter. “Measuring Indirect Effects of Unfair Employer Behavior on Worker Productivity – A Field Experiment.” *Working Paper* (2017).

Hennig-Schmidt, Heike, Abdolkarim Sadrieh, and Bettina Rockenbach. "In search of workers' real effort reciprocity—a field and a laboratory experiment." *Journal of the European Economic Association* 8.4 (2010): 817-837.

Ichino, Andrea, and Giovanni Maggi. "Work environment and individual background: Explaining regional shirking differentials in a large Italian firm." *The Quarterly Journal of Economics* 115.3 (2000): 1057-1090.

Katz, Lawrence F., and Alan B. Krueger. “The rise and nature of alternative work arrangements in the United States, 1995-2015” *ILR Review* 72.2 (2019): 382-416.

Khalmetski, Kiryl, Axel Oxenfels, and Peter Werner. “Surprising gifts: Theory and laboratory evidence.” Journal of Economic Theory 159 (2015): 163-208.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. "The currency of reciprocity: Gift exchange in the workplace." *The American Economic Review* 102.4 (2012): 1644-1662.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. "Do wage cuts damage work morale? Evidence from a natural field experiment." *Journal of the European Economic Association* 11.4 (2013): 853-870.

Landers, Renée M., James B. Rebitzer and Lowell J. Taylor. “Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms.” *The American Economic Review* 86.3 (1996): 329-348.

Leclerc, France, Bernd H. Schmitt, and Laurette Dube. "Waiting time and decision making: Is time like money?" *Journal of Consumer Research* 22.1 (1995): 110-119.

Lourenço, Sofia M. "Monetary Incentives, Feedback, and Recognition – Complements or Substitutes? Evidence from a Field Experiment in a Retail Service Company." *The Accounting Review* 91.1 (2016): 279-297.

Macera, Rosario and Vera L. te Velde. "On the Power of Surprising versus Anticipated Gifts in the Workplace." *Working Paper* (2018).

Malachowski, Dan. "Wasted Time at Work Costing Companies Billions." (2005) Retrieved from https://bxgi.com/content/wasted-time-at-work.pdf

Mas, Alexandre, and Amanda Pallais. "Valuing Alternative Work Arrangements." *The American Economic Review* 107.12 (2017): 3722-3759.

Mas, Alexandre, and Amanda Pallais. "Labor Supply and the Value of Non-work Time: Experimental Estimates from the Field." *The American Economic Review: Insights* 1.1 (2019): 111-126.

Mednick, Sara C, Ken Nakayama, Jose L. Cantero, Mercedes Atienza, Alicia A. Levin, Neha Pathak, and Robert Stickgold. "The restorative effect of naps on perceptual deterioration." *Nature Neuroscience* (2002) 5:677-681.

Merchant, Kenneth A., and Wim A. Van der Stede. „Management Control Systems – Performance Measurement, Evaluation and Incentives." Harlow, UK: Pearson (2017).

Neckermann, Susanne, and Michael Kosfeld. „Getting More Work for Nothing? Symbolic Awards and Worker Performance." *American Economic Journal: Microeconomics* 3 (2011): 86-99.

Neckermann, Susanne, and Xiaolan Yang. "Understanding the (unexpected) consequences of unexpected recognition." *Journal of Economic Behavior & Organization* 135 (2017): 131-142.

Noussair, Charles N., and Jan Stoop. "Time as a medium of reward in three social preference experiments." *Experimental Economics* 18.3 (2015): 442-456.

Oswald, Andrew J., Eugenio Proto, and Daniel Sgroi. "Happiness and Productivity." *Journal of Labor Economics* 33.4 (2015): 789-822.

Park, Sangyoon. "Socializing at Work: Evidence from a Field Experiment with Manufacturing Workers." *American Economic Journal: Applied Economics* 11.3 (2019): 424-55.

Pencavel, John. "The Productivity of Working Hours*." The Economic Journal* 125.589 (2015): 2052-2076.

Prendergast, Canice. "The provision of incentives in firms." *Journal of Economic Literature* 37.1 (1999): 7-63.

Prescott, Edward C. "Why do Americans work so much more than Europeans?" *Quarterly Review Federal Reserve Bank of Minneapolis* (2004):2-13.

Rogers, Todd, and Erin Frey. "Changing Behavior Beyond the Here and Now." *The Wiley Blackwell Handbook of Judgment and Decision Making* (2015). John Wiley & Sons, Chichester, UK

Sandvik, Jason, Richard Saouma, Nathan Seegert, and Christopher Stanton. "Workplace Knowledge Flows." *Working Paper* (2019).

Sherman, Eliot L. “Discretionary Remote Working Helps Mothers Without Harming Non-mothers: Evidence from a Field Experiment.” *Management Science* (2019)

Sliwka, Dirk, and Peter Werner. "Wage Increases and the Dynamics of Reciprocity." *Journal of Labor Economics* 35.2 (2017): 299-344.

Twenge, Jean M., Stacy M. Campbell, Brian J. Hoffman, and Charles E. Lance. “Generational Differences in Work Values: Leisure and Extrinsic Values Increasing, Social and Intrinsic Values Decreasing.” *Journal of Management* 36.5 (2010): 1117-1142.

Whillans, Ashley V., Elizabeth W. Dunn, Paul Smeets, Rene Bekkers, and Michael I. Norton. “Buying time promotes happiness.” *Proceedings of the National Academy of Sciences* 114.32 (2017): 8523-8527.

Wiswall, Matthew, and Basit Zafar. “Preference for the Workplace, Investment in Human Capital, and Gender.” *The Quarterly Journal of Economics* 133.1 (2018): 457-507.

WoldatWork. “Incentive Pay Practices: Privately Held Companies.” (2016) Retrieved from https://www.worldatwork.org/adimLink?id=79934

4.9 Appendix to Chapter 4

4.9.1 Main Hypothesis using a Reciprocity-Based Utility Function

Consider an agent with a quasi-linear utility function similar to section 3:

$$U = u_L(H + H_G + J) + m + m_G - c(J) + \alpha_H(1 - J)H_G + \alpha_m(1 - J)m_G$$

With $u_L' > 0, u_L'' < 0, c' > 0, c'' > 0, J \in [0,1]$. u_L denotes utility from leisure. Leisure can be consumed at home *(H)* and on-the-job *(J)*. The regular working time is normalized at 1 such that the agent's actual working time is *(1-J)*. The second part of the equation represents utility from money m and costs of consuming on-the-job leisure *c(J)* (e.g., potential sanctions when getting caught or deviations from work norms).

Compared to the toy model in section 3, this model adds H_G, an explicit unconditional gift payment in the form of leisure time at home, and m_G, an explicit unconditional gift payment in form of money. The last two terms illustrate a possible gift-exchange effect. If the employee receives an unconditional gift of leisure time at home (H_G) or money (m_G), the employee may reciprocate this by lowering on-the-job leisure J (i.e. by increased work *(1-J)*). α is a constant $\in [0, +\infty]$ capturing the extent to which the agent is sensitive to reciprocate a gift in leisure time at home (α_H) or money (α_m).

The first order condition yields $\frac{\partial U}{\partial J} = u_L'(H + H_G + J) - c'(J) - \alpha_H H_G - \alpha_m m_G$

We are again interested in how the optimal level of on-the-job leisure J^* changes with changes in leisure time at home H. To study possible effects of reciprocity, we are now especially interested in the variation with the gift H_G. By the implicit function theorem, the derivative with respect to the gift of leisure time at home is always negative:

$$\frac{dJ^*}{dH_G} = -\frac{u_L''(H+H_g+J) - \alpha_H}{u_L''(H+H_G+J) - c''(J)} < 0. \quad (1)$$

By the implicit function theorem, the derivative with respect to the gift of money is also always negative:

$$\frac{dJ^*}{dm_G} = -\frac{-\alpha_m}{u_L''(H+H_G+J) - c''(J)} < 0. \quad (2)$$

Comparing the two derivatives (1) and (2) it is now possible to derive a condition under which an unconditional payment of leisure time at home (H_G) will lead to a greater reduction of on-the-job leisure than an unconditional payment of money (m_G) given the reciprocity-based utility function.[37]

$$\frac{dJ^*}{dH_G} > \frac{dJ^*}{dm_G}$$

$$-\frac{u_L''(H+H_G+J)-\ \alpha_H}{u_L''(H+H_G+J)-c''(J)} > -\frac{-\alpha_m}{u_L''(H+H_G+J)-c''(J)}$$

$$\Leftrightarrow u_L''(H+H_G+J)-\ \alpha_H < \ -\alpha_m$$

$$\Leftrightarrow u_L''(H+H_G+J) < \ \alpha_H -\ \alpha_m$$

Since $u_L''(H+H_G+J)$ is negative by definition and α_H and α_m are positive, for the condition to be true, it has to hold that $\alpha_H > \ \alpha_m$. In other words, the sensitivity to which an agent is willing to reciprocate additional leisure time at home has to be greater than the sensitivity to which an agent is willing to reciprocate additional money. In case $\alpha_m > \ \alpha_H$ the concavity of the utility function $u_L\ (H+H_G+J)$ has to be large enough to let leisure time at home be more effective than money in reducing on-the-job leisure.

[37] Note that, for simplicity, I assume here that time and money are measured in the same units.

4.9.2 Figures and Tables

Table 4.3A – Subject and Session Characteristics of Main Treatments

	(1) Baseline	(2) MoneyGift	(3) TimeGift25	(4) All
Age	24.42 (5.85)	23.17 (2.83)	23.45 (3.26)	23.69 (4.23)
Male	0.47 (0.50)	0.36 (0.48)	0.35 (0.48)	0.39 (0.49)
Economics or Business Student	0.44 (0.50)	0.57 (0.50)	0.38 (0.49)	0.46 (0.50)
BDM	7.81 (4.89)	6.50 (3.55)	7.15 (5.86)	7.16 (4.83)
Ability	16.86 (5.96)	15.83 (6.29)	15.65 (5.10)	16.13 (5.79)
Total Sessions	6	8	8	24
Number of Students	43	42	40	125

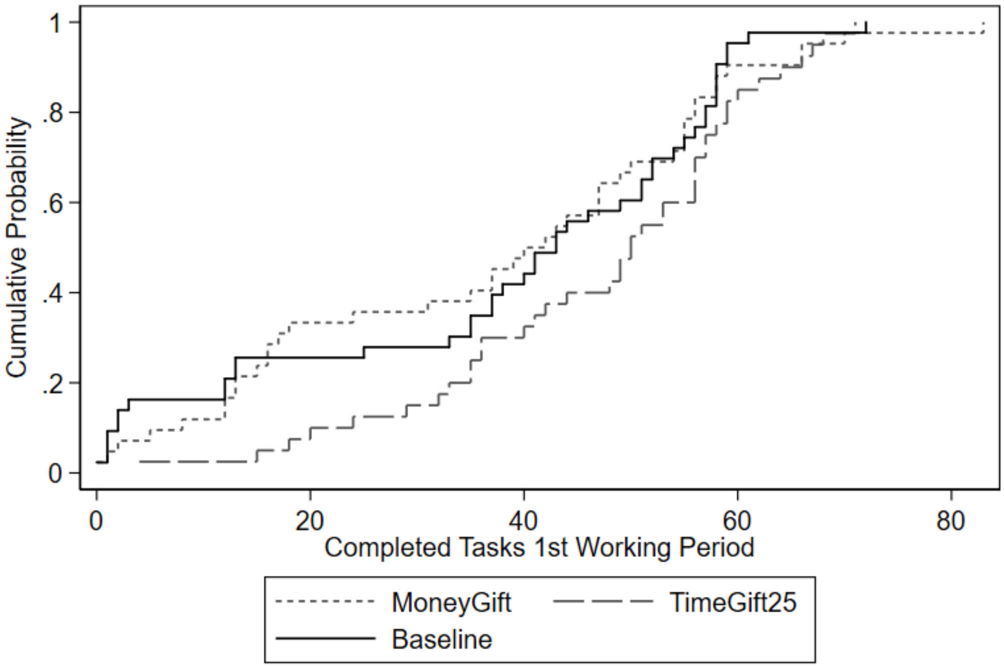
Note: The table displays means and standard deviations (in parentheses) of important subject characteristics in the main treatments.

Table 4.4A – Descriptive Statistics of Main Treatments

	(1) Baseline	(2) MoneyGift	(3) TimeGift25
Total Working Time per Subject	60	60	35
Total Payment per Subject	8	14	8
Completed Tasks 1st Period	37.56 (21.24)	36.93 (21.79)	46.48 (15.91)
Seconds Internet 1st Period	488.93 (660.15)	551.81 (633.90)	224.78 (391.48)
Time Needed per Task 1st Period	39.89 (15.18)	35.63 (8.30)	35.55 (8.27)
Completed Tasks 1st & 2nd Period	77.26 (44.21)	64.62 (43.57)	53.35 (17.88)
Seconds Internet 1st & 2nd Period	1054.95 (1299.67)	1442.28 (1269.01)	290.20 (455.71)
Completed Tasks 1st & 2nd Period per Working Minute	1.29 (0.74)	1.08 (0.73)	1.52 (0.51)
Seconds Internet 1st & 2nd Period per Working Minute	17.58 (21.66)	24.04 (21.15)	8.29 (13.02)
N of subjects	43	42	40

Note: The table reports the means of the important outcomes of the main treatments. Standard deviations are displayed in parentheses.

Figure 4.7A – Cumulative Distribution Function



Note: The figure displays the cumulative distribution function of the key outcome (completed task in the first working period) in the main treatments.

Table 4.5A – Main Treatment Effects – First Working Period

	(1) ln (Total Tasks)	(2) ln (Total Tasks)	(3) ln (Total Internet (in sec.))	(4) ln (Total Internet (in sec.))	(5) ln (Total Tasks)
MoneyGift	0.105	0.182	0.263	-0.049	0.534*
	(0.213)	(0.215)	(0.405)	(0.395)	(0.278)
TimeGift25	0.503**	0.506**	-0.886**	-1.187**	0.608*
	(0.215)	(0.214)	(0.435)	(0.456)	(0.336)
ln (Internet Consumption)					-0.429***
					(0.092)
Constant	3.240***	3.1343**	6.266***	5.739***	4.929***
	(0.150)	(0.648)	(0.292)	(1.533)	(1.201)
Controls	No	Yes	No	Yes	Yes
N of subjects	125	125	125	125	125
Observations	125	125	125	125	125
R^2	0.032	0.074	0.079	0.207	0.376

Note: The table reports coefficients from an OLS regression. The dependent variable is log of the number of completed tasks in the first working period in columns 1, 2, and 5. In columns 3 and 4, the dependent variable is the log of the seconds spent on the Internet in the first working period. Standard errors are in parentheses and clustered at the individual level. All results are compared to the *Baseline* treatment. *ln (Internet Consumption)* is the log of the total amount of time (in seconds) the subjects spent on the Internet in the first working period. *Controls* include age, a gender dummy, a dummy for economics/business students, ability, dummies for the different times of the day, and the WTA. $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

Table 4.6A – Subject and Session Characteristics of Control Treatments

	(1) TimeGift10	(2) TimeGift40	(3) Baseline35	(4) MoneyPfP
Age	23.33	23.15	21.87	23.41
	(4.26)	(3.50)	(2.18)	(3.02)
Male	0.31	0.41	0.38	0.46
	(0.47)	(0.50)	(0.49)	(0.51)
Economics or Business Student	0.44	0.51	0.33	0.64
	(0.50)	(0.51)	(0.48)	(0.49)
BDM	6.66	7.43	8.04	7.57
	(3.33)	(4.21)	(4.06)	(5.17)
Ability	15.10	15.29	16.85	16.26
	(5.62)	(5.05)	(4.81)	(5.91)
Total Sessions	4	4	3	8
Number of Students	43	41	39	39

Note: The table displays the means and standard deviations (in parentheses) of important subject characteristics in the main treatments.

Table 4.7A – Descriptive Statistics Control Treatment

	(1) TimeGift10	(2) TimeGift40	(3) Baseline35	(4) MoneyPfP
Total Working Time per Subject	50	20	35	60
Total Payment per Subject	8	8	8	13.85 (1.07)
Completed Tasks 1st Period	50.10 (15.82)	32.24 (10.50)	44.74 (20.03)	58.36 (10.37)
Seconds Internet 1st Period	151.43 (387.89)	113.54 (238.72)	339.42 (516.23)	7.75 (45.74)
Time Needed per Task 1st Period	34.84 (7.07)	34.25 (6.73)	35.79 (13.30)	31.38 (5.45)
Completed Tasks 1st & 2nd Period	79.41 (26.25)	32.24 (10.50)	53.36 (22.84)	77.38 (22.44)
Seconds Internet 1st & 2nd Period	410.00 (692.41)	113.54 (238.72)	449.59 (798.44)	1102.30 (703.76)
Completed Tasks 1st & 2nd Period per Working Minute	1.59 (0.52)	1.61 (0.52)	1.52 (0.65)	1.29 (0.37)
Seconds Internet 1st & 2nd Period per Working Minute	8.2 (13.85)	5.68 (11.94)	12.85 (22.81)	18.37 (11.73)
N of subjects	39	41	39	39

Note: The table reports the means of the important outcomes of the control treatments. Standard deviations are displayed in parentheses.

Table 4.8A – The Role of Reciprocity

	(1) Tasks per Minute of Working Time	(2) Tasks per Minute of Working Time	(3) Tasks per Minute of Working Time	(4) Tasks per Minute of Working Time
MoneyGift	-0.015	0.001	-0.095	-0.067
	(0.154)	(0.153)	(0.200)	(0.194)
TimeGift10	0.421***	0.462***	0.284	0.344*
	(0.136)	(0.151)	(0.184)	(0.189)
TimeGift25	0.301**	0.295**	0.048	0.071
	(0.176)	(0.138)	(0.186)	(0.187)
TimeGift40	0.357***	0.417***	0.020	0.112
	(0.133)	(0.144)	(0.199)	(0.199)
Reciprocity	-0.173	-0.169		
	(0.136)	(0.132)		
MoneyGift x Reciprocity	0.172	0.166		
	(0.177)	(0.171)		
TimeGift10 x Reciprocity	0.169	0.163		
	(0.149)	(0.148)		
TimeGift25 x Reciprocity	0.243	0.221		
	(0.178)	(0.182)		
TimeGift40 x Reciprocity	0.312*	0.281*		
	(0.160)	(0.164)		
Above Median Reciprocity			-0.393*	-0.377*
			(0.213)	(0.208)
MoneyGift x Above Median Reciprocity			0.188	0.170
			(0.309)	(0.298)
TimeGift10 x Above Median Reciprocity			0.283	0.253
			(0.271)	(0.270)
TimeGift25 x Above Median Reciprocity			0.523*	0.472*
			(0.272)	(0.272)
TimeGift40 x Above Median Reciprocity			0.660**	0.607**
			(0.273)	(0.278)
Constant	1.240***	1.521*	1.436***	1.659***
	(0.160)	(0.371)	(0.136)	(0.359)
Controls	No	Yes	No	Yes
N of subjects	205	205	205	205
Observations	205	205	205	205
R^2	0.112	0.142	0.122	0.152

Note: The table reports coefficients from an OLS regression. The dependent variable is the number of completed tasks per working minute in the first working period. Columns 1 and 2 represent different standardized interactions with the treatment variables. *Reciprocity* is the standardized average investment subjects state that they would return in a hypothetical post-experimental Investment Game. *Above Median Reciprocity* is a dummy variable indicating whether the subjects' reciprocity (measured by the Investment Game) is above or below the median of the subject pool studied in this table. Standard errors are in parentheses and clustered at the individual level. Everything is compared to the *Baseline* treatment. *Controls* include age, a gender dummy, a dummy for economics/business students, ability, dummies for the different times of a day, and the WTA. $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

4.9.6 Survey Results

Table 4.9A – General Evaluation of Time and Money Gifts

	All Managers 1st	Time Survey 1st	Money Survey 1st	Time Survey 2nd	Money Survey 2nd
I think a "Leisure-Bonus" is a new idea.	3.89 (2.07)	3.66 (2.10)	4.11 (2.02)		
I think a "Leisure-Bonus" is a good idea.	5.35 (1.82)	5.24 (1.89)	5.47 (1.76)	5.54 (1.54)	5.63 (1.44)
I think a "Money-Bonus" is a new idea.	1.72 (1.38)	1.56 (1.19)	1.88 (1.53)		
I think a "Money-Bonus" is a good idea.	4.62 (1.64)	4.62 (1.66)	4.63 (1.63)	4.28 (1.59)	4.01 (1.33)
Due to the "Leisure-Bonus", rewarded employees spent less of their working time on leisure activities.	3.49 (1.78)	3.40 (1.72)	3.58 (1.83)	4.93 (1.49)	4.82 (1.50)
Due to the "Money-Bonus", rewarded employees spent less of their working time on leisure activities.	2.27 (1.34)	2.18 (1.23)	2.36 (1.43)	2.75 (1.23)	3.01 (1.14)
Employees/colleagues spend much of their working time on "leisure at work".	3.90 (1.27)	3.72 (1.21)	4.07 (1.32)		
Employees/colleagues spend much of their working time on private use of the Internet.	3.65 (1.41)	3.37 (1.44)	3.90 (1.34)		
N	141	68	73	68	73

Note: The table displays the mean agreement level (1=no agreement, 7=strong agreement) to various statements. Standard deviations are displayed in parentheses. The table compares the mean agreement level of the managers in the time survey (N=68) and the managers in the money survey (N=73) in a between subject design. It further compares agreement at the beginning of the questionnaire (1st) as well as at the end of the questionnaire after managers were confronted with the experimental design of this study (2nd).

Table 4.10A – Different Effects of Time and Money Gifts

For the rewarded employees/colleagues …	Time Survey	Money Survey	*MWU p-value*
…stress would increase.	2.96 (1.49)	2.98 (1.56)	0.9513
…efficiency would increase.	4.54 (1.50)	3.40 (1.38)	0.0000
…privately used time in the workplace would be reduced.	3.75 (1.60)	2.82 (1.44)	0.0005
…the perceived personal appreciation by the employer would increase.	5.63 (1.21)	4.89 (1.44)	0.0001
… health would improve.	4.88 (1.42)	2.52 (1.44)	0.0000
…wasted working hours would increase.	2.91 (1.50)	2.62 (1.34)	0.2974
… job satisfaction would increase.	5.57 (1.22)	4.37 (1.18)	0.0000
…commitment to the company would increase.	5.35 (1.27)	4.51 (1.17)	0.0000
…a potential positive effect would be greatest if they were to carry out office work.	3.46 (1.75)	3 (1.54)	0.1250
… a potential positive effect would be greatest if they enjoy their work.	4.78 (1.73)	5.23 (1.49)	0.1319
N	68	73	

Note: The table displays the mean agreement level (1=no agreement, 7=strong agreement) to various statements. Standard deviations are displayed in parentheses. The table compares the mean agreement level of the managers in the time survey (N=68) and the managers in the money survey (N=73) in a between subject design.

Table 4.11A – Effects of Time Gifts on Work Environment

A "Leisure-Bonus" leads to….	All Managers	*Wilcoxon Signed Rank-p-value – Difference to neutral response (4)*
… capacities for hiring new employees/colleagues.	2.63 (1.68)	0.0000
… a lack of understanding in the team.	3.55 (1.61)	0.0008
… savings in office-, material- and other labor costs.	2.31 (1.44)	0.0000
… lost productivity of the rewarded employee/colleague.	2.40 (1.47)	0.0000
… a better focus of the manager on the rest of the team.	2.57 (1.40)	0.0000
… a positive improvement in the mood of the team.	4.48 (1.45)	0.0000
… an administratively simplified payout compared to a financial bonus.	4.25 (2.04)	0.1737
… no deterioration in the completion of projects.	4.93 (1.61)	0.0000
… a better comparability of the bonus between employees (since for the money bonus there are different monthly incomes and for the leisure bonus there is the same weekly working time).	4.72 (1.76)	0.0000
…operational difficulties.	3.73 (1.56)	0.0523
N	141	

Note: The table displays the mean agreement level (1=no agreement, 7=strong agreement) to various statements. Standard deviations are displayed in parentheses. The table compares the mean agreement level of all managers (N=141) as these statements were provided to both groups (time and money).

4.9.3 Instructions

(originally in German and displayed on the computer screen)

-New Screen-

Welcome

Welcome and thank you very much for participating in today's experiment.

Soon you will receive the instructions for today's experiment. Please read these carefully. In case you have any questions, do not hesitate to notify us by raising your hand. We will be happy to assist you personally.

Throughout the experiment, every form of communication and all activity except participating in the experiment is forbidden. This applies to both verbal and electronic activities. Please turn off your **mobile phone and all other electronic devices** and put them inside your bag. Please put everything else (books, etc.) inside your bag as well and place the bag in front of the wall behind you. Your desk should be completely free from any personal belongings. A violation of this rule will result in exclusion from this and further experiments.

For ease of reading, we abstain from using male and female speech forms simultaneously. All references to persons equally apply to both sexes.

Please click "OK" to continue.

-New Screen-

Task: The slider task

The task you are to complete in this experiment is the **slider task**. This task is required throughout the entire experiment. Therefore, please thoroughly familiarize yourself with the task. In this task, you see **four sliders** on your computer screen. You have to **move all of the sliders into the stated position** by using the computer mouse. The stated position is shown close to the upper border of the screen and applies to all four sliders. If you place all four sliders correctly, you can click the button "**Continue**" in order to work on the next four sliders. These four sliders are then counted as one completed task. Every time you complete one task, the desired position for the following four sliders changes. However, the position of the sliders does not change. Both the number of completed tasks and the time remaining in the working period are shown on your screen.

Additionally, the **opportunity to use the Internet** in order to take a break is part of this slider task. While working on the slider task, you can click the button "**Time Out**". By doing this, you open Internet Explorer, and you can use the Internet. Using the Internet does not involve any disadvantages for you except that the time you spent surfing is deducted from the time remaining for completing the slider task. If you want to proceed with working on the task, simply click the button "**Back to Work**" shown close to the lower border of the screen. The last website you visited remains opened in the background. This means that if you want to use the Internet again, you will return to the exact same position where you left before. You have full anonymity while surfing the Internet. Please note that you can only click the button "Break" in order to use the Internet at the beginning of a task. Once you start with a task, you have to finish it (i.e., move all sliders to the stated position) before you can use the Internet.

Please also note that you can use the Internet, but any other activity (e.g., reading a book you brought) is forbidden and will result in an exclusion from this and all future experiments.

Please click "OK" to proceed to a sample task.

[Instructions for the test stages are omitted.]

- *Screen 1* -

Part 1

Please note that the experiment might **end approx. 30 minutes earlier** than communicated in the invitation email. This means that this experiment would last approx. 90 minutes instead of 2 hours.

In this part of the experiment, we ask you to state the minimum amount of Euros which you are willing to accept in order to, nevertheless, stay in the laboratory for exactly 30 minutes at the end of the experiment (i.e., after approx. 90 minutes) and to continue working on the slide bar task. The task will be absolutely identical to the task you just got to know.

Please click “Continue” to receive more information.

- *Screen 2* -

The following procedure is important for learning your minimum required amount of Euros to stay these 30 minutes. It ensures that you state your true minimum required amount of Euros. Therefore, please thoroughly familiarize yourself with the procedure and notify us if you have any questions.

In a first step, please state the **minimum amount of Euros** that you are willing to accept in order to **work on the slider task for 30 minutes** at the end of the experiment (after approx. 90 minutes). Following this, a random number X will be drawn. This random number lies between 0 and a realistic maximum wage.

One participant will be randomly drawn from the group of participants whose stated amount of Euros **lies below the drawn random number X**. This participant has to stay at her seat and work on the slider task at the end of the experiment (after approx. 90 minutes) for 30 minutes. As compensation for this, the participant **receives a payout in Euros that is equal to the random number X**. Immediately after the experiment, you will be told whether you are the person who was randomly selected or not.

Due to this procedure, it is best for you to state your true minimum amount of Euros!

Here is one example of the described procedure:

Mr. Popeye's stock of spinach is running low. To fill it up, he plans to pull weeds in his parents’ front yard for 6 hours today. His parents propose that they use the procedure presented above and estimate beforehand that Mr. Popeye will probably not demand more than 90 cans of spinach. Thus, they draw a number between 0 and 90. In order for Mr. Popeye to be able to work at his parents' house, he has to come up with the lowest amount of cans that he is willing to accept in exchange for his work. Otherwise, he risks not being able to work at all. However, if he chooses a number that is too low (e.g., 1), the random number will probably be larger than his stated number, and he will most likely able to work for his parents. Yet, in this case, there is a chance that the random number is a 2, which would imply that Mr. Popeye has to work for just two cans of spinach. He says that he wants to receive 30 cans of spinach. His parents randomly draw a 60. Therefore, Mr. Popeye is able to work at his parents' house and, for this, he will receive 60 cans of spinach.

Test questions:

1. Do you have an advantage if you state an amount of Euros that is not equal to your actual lowest amount of Euros?

2. If you are randomly selected, will you receive the amount of Euros you stated?

3. If the random number is larger than your lowest amount of Euros, is there a chance that you will be randomly drawn and have to stay and work?

Please click "Continue" to type in your lowest amount of Euros.

- Screen 3 -

Once again to remind you:

This experiment will **end approx. 30 minutes earlier** than communicated in the invitation email. This means that this experiment will last approx. 90 minutes instead of 2 hours. Now, please state the minimum amount of Euros that you are willing to accept in exchange for staying these exact 30 minutes at the end of the experiment (after approx. 90 minutes) in order to **work on the slide bar task.**

Due to the previously explained procedure, it is best for you to state your actual minimum amount of Euros! You do not have an advantage if you state a number other than your actual lowest amount of Euros.

What is your lowest amount? (in Euro, please use a dot instead of a comma if you want to specify decimal places)

- Screen 4 -

Part 2

The next two parts of today's experiment last **30 minutes each**.

In both parts, you have to work on the presented slider task for 30 minutes. Hence, the next two parts are split as follows:

First working period: 30 minute slider task

Second working period: 30 minute slider task

Please note that these working periods are in no way related to the 30 minutes and the stated lowest amount of Euros from the preceding part. The working periods are thus relevant for all participants as these working periods comprise part of the 90 minutes of the experiment, regardless of the lowest amount of Euros selected. Please click “OK.”

- Screen 5, Baseline -

Compensation

As **compensation** for both working periods, you receive €4.

At the end of the experiment, the compensation and the show up fee will be paid out to you in cash.

Please click “OK” to start the first working period.

- Screen 5, MoneyGift -

Compensation

As **base compensation** for both working periods, you receive €4.

Moreover, you receive an **additional compensation of €6** for **the first working period**. At the end of the experiment, the compensation and the show up fee will be paid out to you in cash.

In the second working period, you do **not** receive an additional compensation.

Please click “OK” to start the first working period.

- Screen 5, TimeGift25 -

Compensation

As **base compensation** for both working periods, you receive €4.

Moreover, **as additional compensation** for the **first working period**, you have the opportunity to leave 25 minutes earlier in the **second working period of the experiment** (another 30 minutes of working on the slider task).

Hence, your working time in the second working period is reduced to 5 minutes, and you can finish the experiment earlier. At the end of the experiment, the compensation and the show up fee will be paid out to you in cash.

In the second working period, you do **not** receive an additional compensation.

Please click “OK” to start the first working period.

- Screen 6, Questionnaire -

Before the experiment, what did you think your compensation for participating in today's experiment would be? (in Euro, please use a dot instead of a comma if you want to specify decimal places)

- *Screen 7, Questionnaire, TimeGift25* -

If you think about your additional compensation in this experiment, would you rather have had the opportunity to earn more money instead of leaving the experiment earlier?

- *Screen 7, Questionnaire, MoneyGift* -

If you think about your additional compensation in this experiment, would you rather have had the opportunity to leave the experiment earlier instead of earning more money?

- *Screen 7, Investment Game* -

Please imagine the following situation:

You and another person (whom you do not know) participate in a study in which you decide upon the distribution of a certain amount of money, thereby determining the pay out of the experiment.

The rules are as follows:

Both participants (you and the other person) receive €20. The other person decides first. She can choose to transfer some of her money to you. She is allowed to transfer any whole number of Euros, i.e., €0, €1, €2 etc. up to €20. Each Euro you receive from the other person will be increased threefold by the team conducting the study and transferred to your bank account.

Therefore, after this round, the other person has €20 **minus** the amount she decided to allocate to you. You have €20 **plus** three times the amount the other person transferred to you.

<u>Now, you have to decide:</u> You have the opportunity to transfer money back to the other person. Depending on the balance of your account, you can transfer back any amount up to €80. This will be the end of the study, and the ultimate balances of the bank accounts will be based on the respective person's final pay out.

The other person's bank account shows €20 **minus** the amount she transferred to you and **plus** the amount you transferred back to her.

You have €20 **plus** three times the amount the other person transferred to you **minus** the amount you transferred back to her.

We would like to know how much you would transfer back to the other person depending on the amount that she transferred to you.

Imagine the other person transfers €5 to your bank account. After the first round, you therefore have 20 + 3*5 = €35. The other person has 20 - 5 = €15. What amount would you transfer back to the other person?

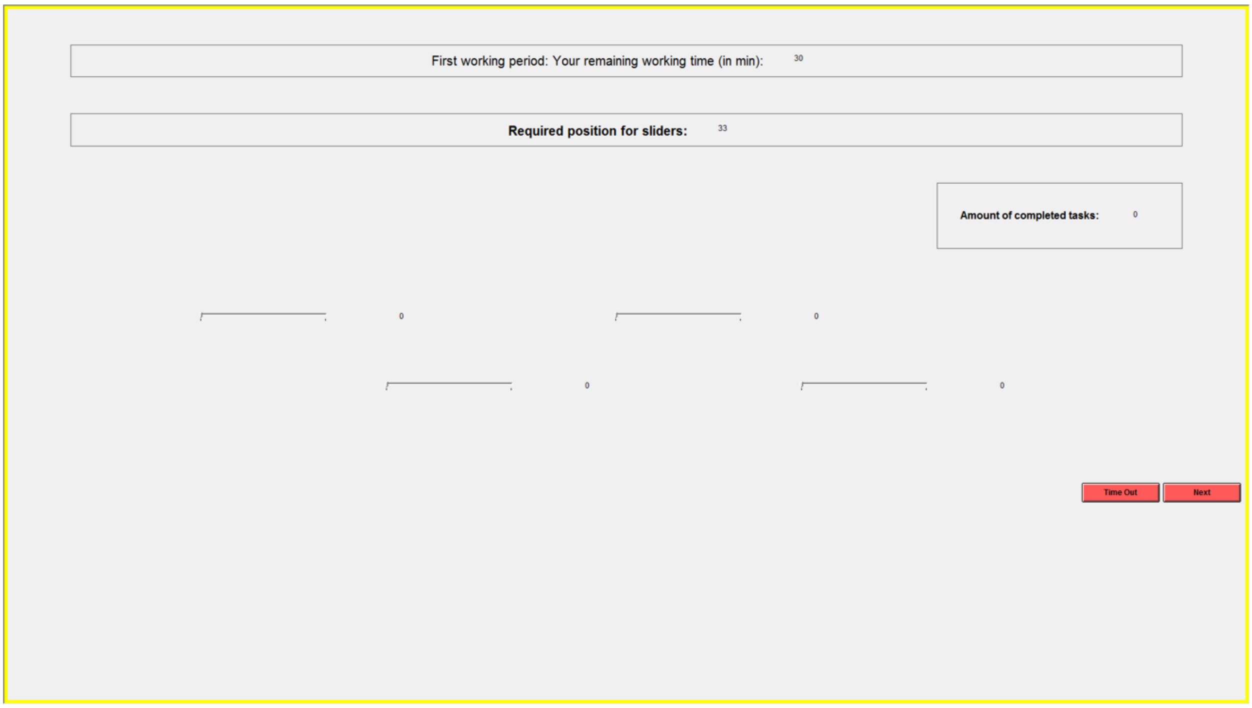
Imagine the other person transfers € 10 to your bank account. After the first round, you therefore have 20 + 3*10 = €50. The other person has 20 - 10 = €10. What amount would you transfer back to the other person?

Imagine the other person transfers € 15 to your bank account. After the first round, you therefore have 20 + 3*15 = €65. The other person has 20 - 15 = €5. What amount would you transfer back to the other person?

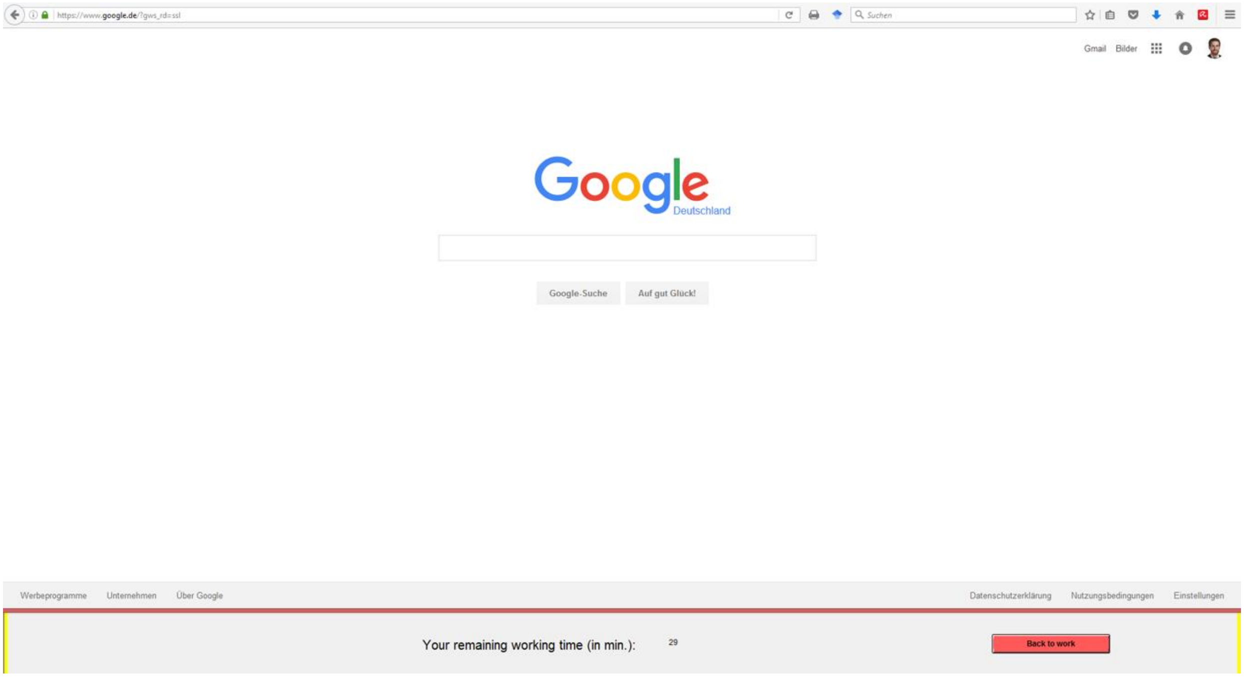
Imagine the other person transfers € 20 to your bank account. After the first round, you therefore have 20 + 3*20 = €80. The other person has 20 - 20 = €0. What amount would you transfer back to the other person?

Applicable to all amounts you state: in Euro, please use a dot instead of a comma if you want to specify decimal places.

4.9.4 Screenshot Working Stage



4.9.5 Screenshot Internet Stage



Eidesstattliche Erklärung

Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen: /

Weitere Personen neben den in der Einleitung der Dissertation aufgeführten Koautorinnen und Koautoren waren an der inhaltlich-materiellen Erstellung der vorliegenden Dissertation nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Dissertation wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Köln, 17.02.2020