I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11859

Performance Pay and Prior Learning: Evidence from a Retail Chain

Kathrin Manthei
Dirk Sliwka
Timo Vogelsang

SEPTEMBER 2018

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11859

Performance Pay and Prior Learning: Evidence from a Retail Chain

Kathrin Manthei
RFH Cologne

Dirk Sliwka
University of Cologne and IZA

Timo Vogelsang
University of Cologne

SEPTEMBER 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Performance Pay and Prior Learning: Evidence from a Retail Chain

We run two field experiments within a large retail chain showing that the effectiveness of performance pay crucially hinges on prior job experience. Introducing sales-based performance pay for district- and later for store-managers, we find negligible average treatment effects. Based on surveys and interviews, we develop a formal model 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, for instance, finding positive treatment effects (only) in stores with low job experience.

JEL Classification: J33, M52, C93

Keywords: performance pay, incentives, learning, experience, insider econometrics, field experiment, randomized control trial (RCT)

Corresponding author:
Dirk Sliwka
University of Cologne
Faculty of Management, Economics and Social Sciences
Albertus-Magnus-Platz
50923 Köln
Germany
E-mail: sliwka@wiso.uni-koeln.de

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[2], 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 large 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 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

[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). More recently, field experiments have started to explore causal effects of performance pay in more complex environments. Hossain and List (2012) study the role of framing bonuses in Chinese high-tech manufacturing. Gibbs et al. (2017) investigate the effect of rewards on innovations in a large technology firm. Delfgaauw et al. (2013, 2014, 2015) run tournament field experiments with a Dutch retailer, and Friebel et al. (2017) study the effect of a team bonus in a German bakery chain.

[2] The experiments were approved by the company's works council, which served as an IRB. The experiments were preregistered with ID AEARCTR-0000961 and AEARCTR-0001758.

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.[3]

In the first experiment, we implemented performance pay based on the average sales per customer for district managers in the fall of 2015. 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) and following the aim of the economist working as a "plumber" (Duflo 2017), 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

[3] 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 and is considered an important 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). For simplicity, we refer to it as the "average sales per customer" in the following.

[4] 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.

habit 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.

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) where knowledge is gradually built up through experience, which leads to concave productivity profiles. We show that when past efforts tend to generate persistent human capital, 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 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 structured into several larger geographical regions that cover Germany and has a rather steep hierarchical structure with relatively small spans of control. 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. The district managers 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. The store managers run a store with about 5-8 full time equivalent employees and are responsible for the daily operation of the store.

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). Ordering decisions made by store managers are based on suggestions generated by a computer system that recommends order quantities using a statistical model. Hence, the store managers are in charge of the execution of operational tasks, such as guaranteeing that shelves are refilled, the store is kept clean, fresh products (fruits, vegetables and bread) are well presented, and that the registers operate efficiently. However, they do have some leeway regarding decisions concerning special placements of goods, temporary price reductions (sales), and product

[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.

orders where they can overwrite the ordering suggestions made by the computer software using potential local knowledge about customer demand.

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.

3 The Experiments

3.1 Experiment I: District Managers

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 A1.1 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]

---

[6] As pre-registered, 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 the 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:

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]

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.

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

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

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

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.

[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] 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. (2016) 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.

[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.

2017, 15 months). Moreover, we provide estimates of the absolute value of the dependent variable. Variations to this are displayed in the Appendix.

Table 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 A1.2 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 A1.3). 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 A1.4 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.

---

[13] Column 3 of Table A1.2 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 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.

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.

3.2 Experiment II: Store Managers

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 A2.1) 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.

3.2.2 Results Experiments II

Again, Table 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 A2.2), monthly treatment effects in Table A2.3. Table A2.4 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.

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 A2.5).[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.

4 Prior Learning and Performance Pay

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 model is a standard moral hazard model with purely transitory efforts. If

[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).

$\phi = 1$, 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 $e_t = \bar{e}$ in each period. Hence,

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

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

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

Hence, we obtain the following result:

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

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

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

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

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

and

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

as by the implicit function theorem

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

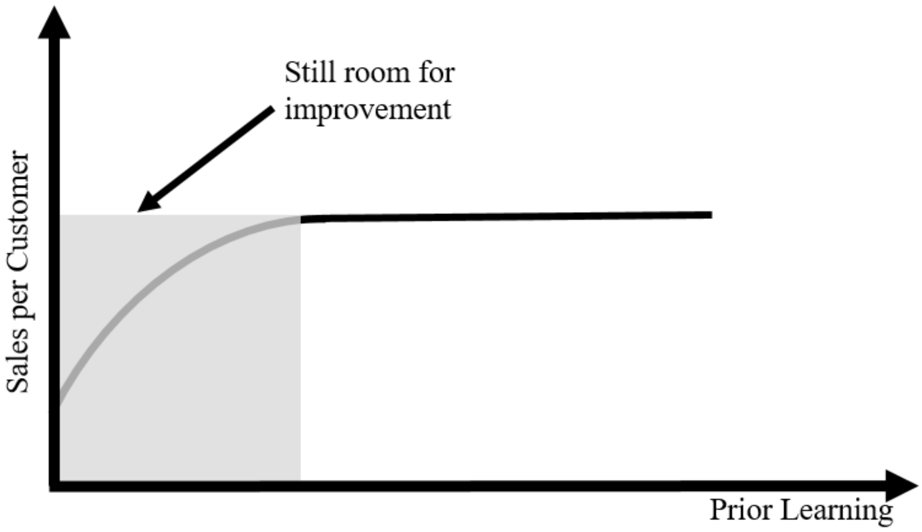
As p_{t-1} is increasing in t the result follows. ■

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.

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 1. The closer a manager is to the beginning of the learning curve (less prior learning), the more room for improvement exists.

Figure 1: 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 A2.5 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 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

[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.

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. 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, Column 2) in both treatment groups.

Table A3.1 in the Appendix reports robustness checks (single difference, longer time periods, trimmed data, log values) and Table A3.2 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: 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 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
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 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 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

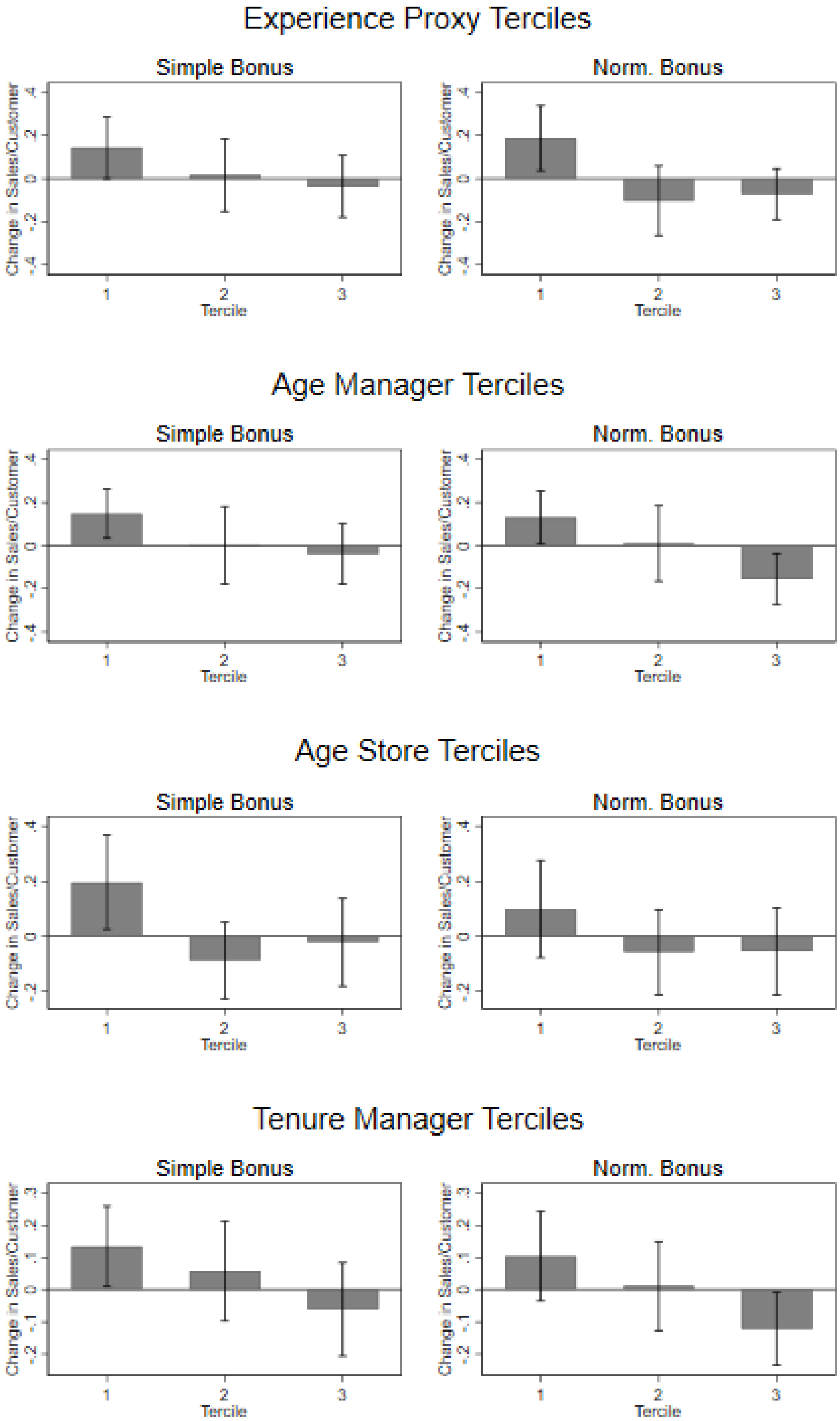
Note: The table reports results from a fixed effects regression with sales per customer on the store level as the dependent variable in different subsamples of the *Experience Proxy*. *Experience Proxy* refers to the mean percentile of a store's age, manager's tenure, and manager's age of the respective 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 Figure 2. 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 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 3.2.3).

Figure 2: 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.

5 Conclusion

We report two large 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. For instance, they can help to understand why firms quite frequently change incentive schemes or the underlying key figures used to measure performance.[33] 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.

REFERENCES

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. "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).

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).

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.

Duflo, Esther. "Richard T. Ely Lecture: The Economist as Plumber." *American Economic Review 197.5* (2017): 1-26.

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

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. "A New Approach to an Age-Old Problem: Solving Externalities by Incenting Workers Directly." *NBER Working Paper No. 22316* (2016).

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.

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.

6 APPENDIX

6.1 Additional Tables Experiment I

Table A1.1: Balancing Table, Experiment I

	(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 A1.2: Robustness Check, Experiment I

	(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. 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 A1.3: Monthly Treatment Effects, Experiment I

	(1) Sales per Customer	(2) 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 A1.4: Other Dependent Variables, Experiment I

	(1) Sales	(2) Customers	(3) Inventory Losses	(4) Mystery Shopping	(5) Ordering Up	(6) Ordering Down	(7) 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.

6.2 Additional Tables Experiment II

Table A2.1: 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)
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 A2.2: Robustness Check, Experiment II

	(1)	(2) Sales per Customer	(3)	(4) 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. 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 A2.3: Monthly Treatment Effects, Experiment II

	(1) Sales per Customer	(2) Sales per Customer
Treatment Effect	-0.0138	0.0048
Norm. Bonus 1st Month	(0.0634)	(0.0524)
Treatment Effect	-0.0184	-0.0142
Norm. Bonus 2nd Month	(0.0492)	(0.0653)
Treatment Effect	-0.0427	-0.0203
Norm. Bonus 3rd Month	(0.0396)	(0.0466)
Treatment Effect	0.100**	0.0978*
Simple Bonus 1st Month	(0.0485)	(0.0565)
Treatment Effect	0.00786	0.0176
Simple Bonus 2nd Month	(0.0468)	(0.0842)
Treatment Effect	0.0166	-0.0134
Simple Bonus 3rd Month	(0.0764)	(0.0599)
Time FE	Yes	Yes
Store FE	Yes	Yes
District Manager FE	No	Yes
Store Manager FE	No	Yes
N Observations	3822	3473
N 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 A2.4: Other Dependent Variables, Experiment II

	(1) Sales	(2) Customers	(3) Inventory Losses	(4) Mystery Shopping	(5) Ordering Up	(6) Ordering Down	(7) 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 Observations	3473	3473	3473	3472	3473	3473	3473
N Stores	294	294	294	294	294	294	294
N 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 A2.5: 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.

6.3 Prior Learning

Table A3.1: 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	1.582	0.302**	0.238*	0.299**	0.0163**
Norm. Bonus	(1.064)	(0.133)	(0.131)	(0.139)	(0.00810)
Treatment Effect	-3.092	-0.590**	-0.437*	-0.590**	-0.0277**
Norm. Bonus x *Experience Proxy*	(1.919)	(0.237)	(0.232)	(0.241)	(0.0132)
Treatment Effect	2.034**	0.331**	0.231*	0.332**	0.0117
Simple Bonus	(0.991)	(0.124)	(0.127)	(0.136)	(0.00767)
Treatment Effect	-3.299*	-0.570**	-0.343	-0.577**	-0.0173
Simple Bonus x *Experience Proxy*	(1.864)	(0.227)	(0.211)	(0.244)	(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. 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 A3.2: Heterogeneous Effects for Position on Learning Curve – Separate Experience Variables

	(1) Sales per Customer	(2) Sales per Customer
Treatment Effect	0.258*	0.312**
Norm. Bonus	(0.129)	(0.139)
Treatment Effect	-0.252	-0.263
Norm. Bonus x *Perc. Tenure Manager*	(0.174)	(0.179)
Treatment Effect	-0.141	-0.133
Norm. Bonus x *Perc. Age Store*	(0.172)	(0.190)
Treatment Effect	-0.130	-0.210
Norm. Bonus x *Perc. Age Manager*	(0.168)	(0.155)
Treatment Effect	0.224*	0.282*
Simple Bonus	(0.130)	(0.143)
Treatment Effect	-0.0930	-0.153
Simple Bonus x *Perc. Tenure Manager*	(0.156)	(0.152)
Treatment Effect	-0.148	-0.136
Simple Bonus x *Perc. Age Store*	(0.142)	(0.174)
Treatment Effect	-0.150	-0.205
Simple Bonus x *Perc. Age Manager*	(0.153)	(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
Observations	3692	3378
N of Stores	284	284
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.

6.4 Instructions (for online Appendix)

6.4.1 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.

Calculation of performance measure average receipt

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 =

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

%-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"

6.4.2 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.

6.4.3 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.

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 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.

<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 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

6.4.4 Monthly Notifications Experiment II

Normalized 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 versus nation</u>

- 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 <u>increase versus nation</u> = %-increase receipt in the current month versus previous year (store)
– %-increase receipt in the current month versus previous year (nation)
= **XXX**

<u>Constant base value</u>

- 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 <u>base value</u> = %-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

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 compared to previous year

- 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.