

Talking about Performance or Paying for it? – A Field Experiment on Performance Reviews and Incentives

Kathrin Manthei, Dirk Sliwka and Timo Vogelsang*

This Version: September, 2021

We investigate the causal effect of performance pay and conversations about performance implementing a 2x2 field experiment in 224 stores of a retail chain. In the *performance pay* treatments, managers receive a bonus which is a simple linear function of the profits achieved above a threshold value. In the *performance review* treatments, managers have to report their activities undertaken to increase profits in regular meetings. We find that while performance pay did not yield significant profit increases, performance review conversations increased profits by about 7%. However, when additionally receiving performance pay, the positive effect of performance reviews vanished. We provide evidence from surveys and meeting protocols that performance pay changes the nature of conversations leading to a stronger self-reliance of store managers which undermines the value of the performance reviews.

Keywords: performance pay, monitoring, feedback, field experiment, management practices.

JEL Classifications: J33, L23, M52, C93

Acknowledgments: We thank Jordi Blanes i Vidal, Robert Dur, Florian Englmaier, Maria Guadalupe, Matthias Heinz, Eva Labro, Simeon Schudy, and Wim Van der Stede. for helpful comments. Moreover, we like to thank participants of 5th IMEBESS in Florence, the 11th MBEES in Maastricht, the Ohlstadt Workshop on Natural Experiments and Controlled Field Studies 2018, the COPE 2018 in Munich, the 4th INSEAD RCT days 2019, and the Advances with Field Experiments 2020, as well as seminar participants at the London School of Economics. Jakob Alfitian, Katharina Arnhold, Sidney Block, Sophia Schneider, Caro Wegener, Julia Schmitz, and Theresa Hitzeman provided outstanding research assistance. We are very grateful for the assistance by the company and their staying power in working with us. No funding was received from the company, no coauthor had a financial relationship with the company, and none of results were corrected. Any errors and all opinions are our own. The University of Cologne did not have an IRB at the time the experiment was carried out. However, the rules of the laboratory were used to enforce an ethical standard. Moreover, the works council approved of all aspects of the experimental design serving as an IRB substitute. The experiment was registered with the ID AEARCTR-0002127. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866.

Kathrin Manthei: Rheinische Fachhochschule Cologne, University of Applied Sciences, Schaevenstrasse 1a/b, 50676 Köln, E-mail: kathrin.manthei@rfh-koeln.de

Dirk Sliwka: University of Cologne, Department of Personnel Economics and Human Resource Management, Albertus-Magnus-Platz, D-50923 Köln, Germany, E-Mail: sliwka@wiso.uni-koeln.de

Timo Vogelsang, Frankfurt School of Finance & Management, Department of Accounting, Adickesallee 32-34, Frankfurt, Germany, E-Mail: t.vogelsang@fs.de, tel. +49 69 154008 – 393

1. Introduction

Influencing employees to act in the interest of employers has been a key focus of the literature in management, accounting, and economics (see, e.g., Prendergast 1999, Sprinkle and Williamson 2006, Gibbons and Roberts 2013, or Lazear 2018 for surveys). Scholars have traditionally stressed the importance of monetary incentives to align employees' behavior with the objectives of employers. Indeed, a number of field experiments and quasi-experiments in firms have shown that financial incentives can have substantial performance effects (see, e.g., Lazear 2000, Banker et al. 2000, Bandiera et al. 2007, Hossain and List 2012, Delfgauw et al. 2013, Lourenço 2016, Friebe et al. 2017).¹

Organizations of course also adopt a wider array of management practices to align employees' actions with the organization's objectives (Bloom and Van Reenen 2007). In particular, the behavior of employees is often guided through simple conversations between superiors and subordinates in which subordinates report their activities. The key aim of this paper is to investigate whether such "performance review" conversations and performance pay are alternative solutions to reduce moral hazard problems. In particular, we (i) study the causal effects of regular conversations *about a specific outcome variable*, (ii) compare the effects of conversations to the effects of performance pay based on the same outcome variable, and (iii) evaluate interaction effects when both practices are introduced at the same time.

From an economic perspective, structured conversations in which employees have to report their activities to raise a specific outcome variable can be viewed as monitoring devices. As, for instance, has been put by Lazear and Oyer (2013, p. 486): "An alternative to financial incentives is to simply monitor workers. If a supervisor can keep close watch over employees, she can ensure that the employee takes the best actions". When employees anticipate that they will have to explain their actions undertaken to foster an objective in regular conversations with their supervisor, this should trigger additional reputational concerns. These reputational concerns through the "craving for appreciation and the desire to impress superiors" (as put by Holmström 2017, p. 1772 in his Nobel lecture) can be a powerful motivator which may reduce

¹ Surveys of the extensive literature on performance pay can be found in Bandeira et al. (2011), List and Rasul (2011), Levitt and Neckermann (2015).

the need to use performance pay.² On top of this, the review conversations may provide feedback to employees helping them to improve their performance.

We conducted a field experiment in a retail chain with 224 store managers.³ Prior to our intervention store managers' performance was mostly evaluated based on store sales. With the intervention the firm introduced a simple profit metric. Store managers were randomly assigned to one of four treatments in a 2x2 design:

- Performance pay
- Performance reviews
- Performance pay & performance reviews
- Control group.

Store managers in the performance pay conditions received bonus payments that were simple linear functions of profits achieved above a threshold value (planned budget). In the performance review conditions, the store managers' supervisors were asked to arrange meetings every second week. In these meetings, store managers had to report their activities to raise profits as well as their intended next steps following a pre-determined protocol. The protocol did not include specific triggers for "motivational talk" (compare Kvaløy et al. 2015 or Antonakis et al. 2021) but focused specifically on the reporting of activities.

To assure that all store managers were well informed about the newly introduced profit metric, each manager received an information package consisting of a one-time online training on the underlying profit measure and information about profit margins of products. To exclude that the results are driven by attention generated for this profit metric through the information package, managers in the control group were also provided with it.⁴ We investigate the hypotheses that (i) the introduction of performance review conversations on a specific objective and (ii) the use of performance pay based on this objective both raise performance. Furthermore, we hypothesized that (iii) both management practices are (imperfect) substitutes, i.e. that the

² Fama (1980) first made the claim that bonus payments may not be necessary when career concerns discipline managers. Holmström (1999) formalized the idea showing that career concerns indeed reduce the moral hazard problem. Gibbons and Murphy (1992) explicitly model the interplay between explicit incentives and career concerns and show that both are (imperfect) substitutes to raise performance.

³ See Harrison and List (2004), Bandiera et al. (2011) or Floyd and List (2016) for recent surveys on field experiments.

⁴ Before the intervention, store performance was mostly assessed by tracking single components of store profits such as sales and inventory losses and store managers did not know profit margins. As a first step, a simple profit metric was introduced in all treatments as an aggregate accounting return measure. The online training and feedback on the profit measure assured that managers were provided with information on profit margins and helped to brush-up the managers' knowledge.

effect of introducing performance pay is weaker when performance reviews are in place and vice-versa.

Our key results are the following: We find that performance pay alone did not raise profits significantly above the level in the control group. But the performance review conversations increased profits substantially by about 7%. Moreover, the positive effect of the performance review conversations vanished when accompanied by performance pay. Hence, asking managers to report their activities to raise a performance objective increases this objective substantially at rather low costs. However, performance pay clearly undermined the value of the review conversations: While we hypothesized that performance pay reduces the *marginal* effect of introducing performance reviews, we found – in contrast to our expectation – that it in fact even reduces the *absolute* effect of this practice.

The result that performance pay does not raise profits above the level of the control group seems surprising but is in line with results reported in Manthei et al. (2021b). In this companion paper we ran a field experiment in a different region of the same company in which we varied whether store managers receive a bonus, the information package, or both. As that study shows both practices, raise profits significantly when applied separately, but the profit growth induced by the combined intervention does not significantly exceed those of the separate interventions. The paper argues that both instruments are to some extent substitutes to generate attention for the underlying profit metric fostered by both. Our results here show that performance reviews are powerful in generating returns above the effect of the information intervention and, moreover, that bonuses undermine this effect.

To understand why performance pay was detrimental for the conversations, we collected detailed information on store managers' activities, the number of meetings conducted, the content of the conversations, and post-experimental survey questions on satisfaction as well as the relationships to district managers. In particular, supervisors conducting the reviews in our field experiment had been asked to write down short protocols noting activities undertaken by the respective store managers, occurring problems, and next steps. Investigating these protocols, we find that the use of performance pay shifted behavior in the review conversations. A key change is that – while there are no significant treatment differences in the number of reported activities and intended next steps – store managers mention occurred problems at a much lower frequency when there is performance pay. In fact, store managers state nearly three times more problems per conversation when there was no bonus. About 60.3% of the store

managers who received a bonus mentioned not a single problem in any of the meetings and this fraction is only 22.2% among those without a bonus. Bonuses thus shifted behavior towards a stronger self-reliance of store managers – and this obviously backfired: When we replicate our main regressions including only observations from stores in which conversations included a discussion of problems, the negative effect of the bonus on the performance effect of review conversations vanishes. Quite strikingly, store managers’ overall assessment of the set of practices provided is more favorable in the group that has the review conversations but received no bonus and the perceived quality of feedback is also higher. We also provide a rationale for this finding in a formal economic model. We show that bonuses based on individual performance rationally lead to a stronger self-reliance when employees fear that activities that involve the supervisor (such as stating a problem or asking for help) may be interpreted as being selfishly driven rather than by genuine interest in solving the issue.

Our paper contributes to the recently growing literature evaluating the impact of monetary and non-monetary management practices on performance in field experiments. Positive performance effects of monitoring activities have, for instance, been found by Nagin et al. (2002) for call center workers and Gosnell et al. (2020) for airline pilots. Banker et al. (2018) investigate how the prior intensity of monitoring is associated with the benefits of introducing performance pay in a field experiment in retailing. They find that the marginal impact of performance pay may be negative in units with high levels of prior monitoring.⁵ Examples of further specific non-monetary practices that have been studied using field experiments include recognition (e.g., Bradler et al. 2016, Lourenço 2016), employee performance rankings/feedback (e.g., Barankay 2012, Eyring and Narayanan 2018, Blader et al. 2020)⁶, upward feedback from employees to supervisors (e.g., Cai and Wang 2020) or facilitating managers’ decision through the provision of information (e.g., Bloom et al. 2013, Hanna et al. 2014, Casas-Arce et al. 2017a, Manthei et al. 2021b).

The results also contribute to recent field experimental work on communication in organizations. Sandvik et al. (2020), for instance, implemented randomly assigned brief

⁵ See also Boly (2011) and Belot and Schröder (2016) for lab and field experiments on monitoring in which sanctions are tied to the result of the monitoring outcome such that the effects of monitoring are not disentangled from the effects of performance pay. Campbell et al. (2011) use data from a casino chain in which each casino could decide about the intensity of the monitoring. They find that tight monitoring leads to strong implicit incentives, which leads to less experimentation and learning.

⁶ The effect of quantitative (rank) feedback on performance has been studied extensively in recent years with rather mixed results. While some field studies find positive effects on performance (e.g. Blanes i Vidal and Nossol 2011, Tran and Zeckhauser 2012, Blader et al. 2020, Eyring et al. 2021) other find negative effects (Barankay 2012, Ashraf et al. 2014, Bursztyn and Jensen 2015, Bradler et al. 2016), no effect (Lourenço 2016) or mixed effects (Casas-Arce et al. 2017b, Eyring and Narayanan 2018).

meetings among coworkers in a firm to discuss sales techniques. They find that such meetings substantially increase average sales revenue. In contrast to our setting, meetings in their paper are organized between employees on the same hierarchical level to foster knowledge exchange. Our setting, however, is concerned with performance review meetings in which managers have to report their activities to their supervisors. Burgess et al. (2019) implement a peer feedback scheme among teachers in 82 high schools where teachers were randomly assigned to the role of observer and observee and find that this improved the performance of both groups. Casas-Arce et al. (2019) show that managerial attention through branch visits in a retail bank increases branch performance after the visit. In Friebel et al. (2021) supervisors in randomly selected supermarkets were asked by the top management to lower employee turnover which indeed significantly reduced the number of quits. Supervisors attributed this in particular to stronger personal interaction with their employees.⁷ Our results provide further support for the power of fostering communication in organizations – but we also find that the use of performance pay affects the nature of communication and can undermine its value.

The results also add to the literature on detrimental effects of incentives (e.g. Gneezy and Rustichini 2000a, 2000b, Bénabou and Tirole 2003, 2006, Fehr and Rockenbach 2003, Fehr and List 2004, Falk and Kosfeld 2006, Sliwka 2007, Ellingsen and Johannesson 2008, Ariely et al. 2009, Christ 2013, Cardinaels and Yin 2015, Alfitian et al. 2021) which has so far mostly shown the existence of such detrimental effects in laboratory experiments rather than in the field within firms.⁸ In our study, performance pay is not detrimental when introduced in isolation, yet becomes detrimental when structured conversations about performance are in place. The reason is that performance pay changes the nature of these conversations and undermines their value leading to a stronger self-reliance of employees.

The results also complement arguments put forward in the literature on biases in subjective performance evaluations of employees. Prendergast (1999), for instance, discusses the argument that bonus payments can change the nature of these biases and that “[...] *many firms now explicitly separate pay setting from subjective evaluations*” (Prendergast 1999, p. 30).⁹ In recent years, many larger firms have indeed revised their procedures to manage

⁷ In a similar vein, we also contribute to the broader empirical literature on the value of supervisors (Bertrand and Schoar 2003, Lazear et al. 2015, Bender et al. 2018, Bandiera et al. 2020, Hoffman and Tadelis (forthcoming)) and on the causal effects of leadership (see, e.g., Kvaløy et al. 2015, Antonakis et al. 2021).

⁸ For recent surveys on the interplay between monetary and non-monetary incentives see e.g. Ellingsen and Johannesson (2007), Gneezy et al. (2011), Rebitzer and Taylor (2011) or Cassar and Meier (2018).

⁹ His argument, however, is different from the mechanism we suggest. He conjectures that poor feedback may be harder to communicate for supervisors once a monetary bonus is attached to the rating. In our setting, the performance reviews do not

employee performance and, in particular, often have reduced the role of individual rewards and focused more on establishing a regular dialogue about performance.¹⁰ Moreover, it has even been claimed that bonuses may undermine open communication and in turn harm performance.¹¹ Our results indicate that such claims are not lacking substance.

The fact that the positive effect of the conversations vanish with performance pay further contributes to the literature on interdependencies between different management practices in organizations (Milgrom and Roberts 1990, 1995, Ichniowski et al. 1997, Ferreira and Otley 2009, Grabner and Moers 2013, or Brynjolfsson and Milgrom 2013 for a recent survey) which argues that the impact of a specific management practice will often depend on the use of other practices as there will be complementarities or substitution effects. We contribute to this literature showing that the use of different management practices that are intended to achieve the same outcome may even be worse than the use of a single practice.

2. The Organization

The company is a nationwide retailer, operating discount supermarkets in Germany consisting of several larger geographical regions.¹² The companies' structure and tasks on the different hierarchical level are visualized in Figure 1.

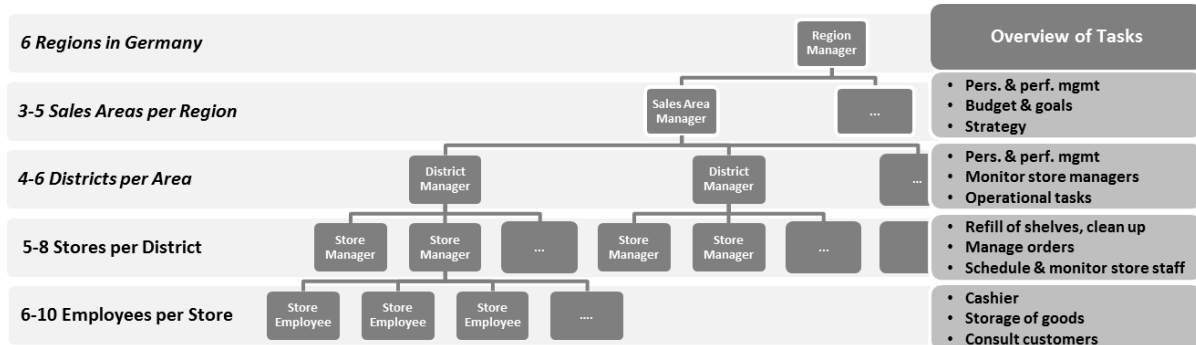
determine the bonus payment, as the bonus is based on objective performance measures and the supervisors have no influence on the size of the bonus.

¹⁰ Cappelli and Tavis (2016) survey the development of performance management practices in larger firms. Examples for larger firms that recently entirely stopped or strongly revised standard annual performance ratings tied to rewards and instead established regular feedback conversations are for instance Adobe, Deloitte, Lear, Microsoft, IBM, Bosch, or SAP.

¹¹ Tom DiDonato, Chief Human Resource Officer of Lear Corporation for instance claims that "*Performance reviews that are tied to compensation [...], discourage straight talk, and too easily become politicized.*" (DiDonato 2014). Uwe Schirmer, Head of HR Policies at Bosch, world's largest auto parts supplier, for instance, claims that "*feedback discussions have become less tactical*" since Bosch has abolished individual performance bonuses in 2015 (Handelsblatt, Nov, 11 2018).

¹² The company is the same as in Manthei et al. (2021a) but the studies are conducted in different regions.

Figure 1: Illustrative Organizational Chart



Each region has a regional top manager (region manager) and sales area managers. The sales area managers supervise about 4-6 district managers. District managers are responsible for 5-8 store managers. A store employs approximately 5-8 (full time equivalent) employees.

Store managers are responsible for the daily routines within the store. They have limited leeway within their operational tasks as discount retailing is generally characterized by highly standardized tasks and processes (for instance, concerning the placement and ordering of products). While a computer system generates recommendations for order quantities of products, store managers can overwrite these suggestions. Moreover, they can decide on special placements of goods within a limited area in the store. Store managers' main duties lie in the execution of daily operational tasks such as keeping the store clean, the presentation of products, the availability of products in the shelves and an efficiently working cash desk (see Table A1 in the Appendix for an overview of tasks). They are also responsible to generate the shift schedule. The main duty of district managers is supervision of store managers, whom they visit approximately twice per week. During these visits, the district managers examine the stores and discuss certain topics with the store manager. These topics are mainly related to the introduction of new product placements within the store and personnel management issues. In cooperation with the sales managers, they are also in charge of the personnel and performance management as well as the budget planning of their districts.

With their store computer, store managers are able to access data on the performance of the store. Specifically, store manager receive an electronic overview on several KPIs (such as sales, customer frequency, personnel expenses, inventory losses etc.) each and every week and

month. They further see their rank for each KPI within the region. District managers are able to access this information for each of their store managers.

All store managers in the respective region receive a small annual bonus based on sales, inventory and a mystery shopping score, which accumulated to on average €233 per store manager in 2017. The design of this existing bonus scheme was such that payments were fixed within brackets so that bonus payments hardly varied over time. While the existing bonus plan provides rather low-powered incentives, store managers have considerable career incentives. Most district managers are former store managers and approximately 5% of store managers are promoted to become a district manager in a given year. The wage differential between district and store managers is quite substantial as the wage of district managers is nearly twice as large as that of store managers (gross monthly salaries are about €6000 for district and €3000 for store managers).

The specific characteristics of the stores and managers within our region of focus are summarized in Table A2.

3. The Experiment

3.1. Design and Procedures

We introduced performance pay and performance reviews for store managers, implementing a 2x2 factorial experimental design over three months (April – June 2017). Prior to the intervention, store managers were not systematically trained to work with store profits and were mostly concerned with managing sales.¹³ The key aim of the intervention was to get store managers to focus more on profits, thus broadly taking into account the effects of their actions on both sales and the respective costs. The key performance metric for this intervention is a simplified form of the store's profits

$$\textit{Profit} = \textit{Net Sales} - \textit{Cost of Goods Sold} - \textit{Staff Costs} - \textit{Inventory Losses}$$

which covers all key elements of company performance a store manager can influence.¹⁴

¹³ See Manthei et al. (2021a) for an earlier field experiment within the same organization but among different store managers, where a sales based rather than profit based bonus was evaluated. Importantly, the study finds close to zero average treatment effects due to a strong focus on sales increases prior to the intervention, limiting the scope for further sales growth. This insight led to the introduction of store profits as a relevant performance metric which allows a broader scope of actions that store managers can undertake to increase performance.

¹⁴ It excludes, for example, rent payments or costs of renovations on which store managers have no influence.

We randomly assigned each district within the region to one of the following treatments: *BONUS*, *REVIEW*, *BONUS&REVIEW*, or the *CONTROL* group.¹⁵ Store and district managers did not know that this was part of an experiment and we thus maintained a natural environment. Table 1 summarizes the treatments.¹⁶

Table 1: Treatments

	Review	No Review
Bonus	N=63	N=51
No Bonus	N=50	N=60

Importantly, prior to the intervention all store managers received an information package about the profit metric. To ensure that treatment effects are not merely driven by creating attention for the new performance metric (compare e.g. Manthei et al. 2021b) this information package was also provided to managers in the control group. The package included an online training consisting of a video and a quiz about possible ways to increase stores profits and novel information about the relative profit margin (e.g. $(sales\ price - procurement\ price)/sales\ price$) of each product. For this, all products are ranked according to their margin and then divided into five equal sized groups and named SP1 (highest margin) to SP5 (lowest margin). All store managers further received a monthly report about the development of profits (and its components) apart from the possible monthly bonus notifications. From the end of March onwards, store managers had access to the online training.¹⁷

3.2. The Treatments

In the following we describe the design of the treatments in more detail.

¹⁵ The experiment was preregistered under AEARCTR-0002128. Note that we initially registered two regions but randomized within each region. However, in one region (North-West Germany) the regional manager told us already in the early weeks of the interventions that higher-level management did not back the project due to many refurbishments in the region and difficult external market influences during that time. Due to this, district managers did not regularly hold the conversations with the store managers (average number of 3.5 conversations per store in the focus region, 2.4 conversations in the other region, MWU $p < 0.001$). Moreover, they did significantly fewer first conversations in the first two weeks (90% conversations in the focus region, 52.48% conversations in the other region, Signed Rank Test $p < 0.001$). Importantly, bonus payments were also delayed by the regional manager and corrected ex-post. However, due to randomization within each region not focusing on this region does not bias our estimates.

¹⁶ Differences in the sample size per treatment occur due to our randomization on the district level.

¹⁷ Importantly, it was not possible for store managers to infer from the video and the quiz that they are part of a study run by a university.

BONUS

In each of the three months from April to June 2017, store managers in this group were eligible to receive a bonus according to the following formula:

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

Store managers received €0.05 for every €1 profit above a threshold of 80% of the planned budget value. Based on a prediction algorithm, the planned budget is calculated by the accounting department in the beginning of the year. The calculation is based on previous performance as well as other influencing factors such as renovations, and opening or closing of stores by competitors. Store and district managers have no possibility to influence this value. Importantly, store managers could expect that the bonus provides substantial additional earnings. One month before the experiment 15 out of the 51 store managers in this treatment were below the planned profit budget. For the store managers in this treatment a profit increase by 10 percentage points above the threshold would imply on average about €109 higher earnings.

Accumulated bonuses were paid out after the three months of the experiment with the store managers' salary. Store managers were informed with personalized letters each month from April to June 2017.¹⁸ The letter reported the achieved profit of the store and all of its components of the previous month. Moreover, the initially planned budget values were also provided. Additionally, store managers received feedback about the bonus for the respective month.¹⁹

REVIEW

In this treatment, store managers had systematic biweekly²⁰ conversations with their supervisors (district managers) about their actions to increase store profits as measured by the

¹⁸ More precisely, due to a delay in calculating staff costs, the data were always delayed by one month. Hence, for instance, by the end of May letters were sent out with the calculations for April. However, although the official letters were delayed, store managers received electronic performance feedback directly at the end of the month. Hence, they were able to infer their performance. Furthermore, store managers were used to time lags for certain metrics which are common in the company, e.g. the calculation of inventory losses is subject to delays for procedural reasons.

¹⁹ Note again, that all store managers in this region also received an annual bonus for sales, inventory and a mystery shopping score which accumulated to on average €233 per store manager in 2017. The design of this existing bonus scheme was such that payments were fixed within brackets so that bonus payments hardly varied over time.

²⁰ The biweekly time frame was chosen to allow store managers sufficient time to implement actions but still have a rather tight monitoring scheme. As monitoring here was introduced to motivate employees and not for selection purposes drawbacks from too tight monitoring as for instance analyzed by Ichino and Mühlheusser (2008) (who show that monitoring can harm optimal selection as monitored agents then hide their true type when monitoring is too intensive) should not apply.

metric described in the above. It is important to note that the aim of the experiment is therefore not to assess the role of broad feedback conversations with supervisors, but to evaluate the specific role of performance review conversations about activities undertaken *to raise a specific performance objective*. Therefore, the intervention does not include triggers for “motivational conversations” but focuses on a review of performance in the narrow sense of the word.²¹

District managers were provided with a “Conversation Guide” with three specific questions to be discussed during the conversations which all induce store managers to report their activities to raise profits. These questions were:

- (a) “What did the store manager do to increase profits?”
- (b) “Which problems occurred?”
- (c) “What does the store manager plan to do before the next meeting?”

District managers were asked by the Human Resource (HR)-department to write a protocol documenting the responses to these three questions and send it back to the HR office. The HR office then forwarded them to us. District managers received emails every two weeks to remind them to have the conversations. They were also not aware that they were part of an experimental study.²²

BONUS&REVIEW

This treatment is a combination of individual monetary performance pay and the biweekly conversations with the district managers. The characteristics are as described above.

CONTROL

All conditions for the CONTROL group remained unchanged. However, for reasons of symmetry as already stated above we provided the video, the quiz and the product margin information also to store managers in the control group to avoid that the results are driven by attention effects.

²¹ Kvaløy et al. (2015) or Antonakis et al. (2021) show that for instance motivational speeches can have a substantial impact on performance.

²² We use the following phrase in the introductory letter “We would like you to have an intensive, personal conversation with your store manager each and every second week”. Store managers were informed at the beginning of the treatment phase that their supervisors will meet them every second week for the review conversations.

3.3. Implementation

We use a stratified randomization depending on a prediction of the district profits for the first treatment month (see, e.g., Athey and Imbens 2017). To construct the stratification groups, we predict profits for the district in April 2017 using one year of past data through January 2017 with a simple fixed effects model ($\text{Adj. } R^2=0.9241$).²³ We then randomly assigned the treatments within groups of four with similar predicted values. This aims to reduce the standard error in our main variable of interest. Randomization was conducted at the district level in order to avoid possible spillover and contamination effects between different stores as store managers communicate (infrequently) within districts but very rarely across districts. We provide a balancing table in Appendix A2.²⁴

Personalized letters were sent to the store managers' home addresses in the last week of March to inform them about the changes. The letters were signed by the region manager and the regional HR manager, and sent from the company's post office with no indication that a university is involved or an ongoing experiment.²⁵ We also ran two online surveys with store and district managers before and after the experiments. The surveys contained, for instance, questions about the daily work focus of managers and their job satisfaction. The second survey additionally included open questions about which actions the managers actually undertook in the past to increase store profits. Again, personalized letters were directed to the store managers' home addresses in February 2017 as well as in the last week of June 2017. These letters were sent from the university as an independent research organization to guarantee anonymity to the managers. Moreover, they were introduced as a general survey on satisfaction and work behavior. There is thus no obvious connection between the surveys and the interventions.

3.4. Hypotheses

Our pre-registered key hypotheses for the field experiment were that (i) bonus payments for profit increases and (ii) performance review conversations on profit increases both raise profits, and (iii) both instruments are (partial) substitutes. The first hypothesis directly follows from the rich literature on moral hazard: if store managers receive a bonus based on store profits,

²³ We had to randomize three months in advance as the data on profits, as explained above, were accessible with a delay of one month.

²⁴ The small imbalances that are visible in the balancing table are idiosyncratic. Moreover, they are time invariant and should be controlled for in fixed effects regression. However, we further control for the imbalanced variables in Table A3 in the Appendix to show robustness of our results.

²⁵ Exemplary letters are provided in the Appendix 7.4.

they should have stronger incentives to work harder to raise these profits. The second hypothesis is based on the fact that review meetings implemented induce a closer monitoring of the agents' actions. In the respective treatment groups store managers were required to report their activities to raise profits. In turn, they should face psychological costs when underperforming or benefit from reputational gains when demonstrating high efforts and this should raise their incentives to perform. When performance reviews provide a less noisy signal of store managers' efforts they should, for instance, directly increase career concerns incentives as in Holmström's (1999) model. As explained in section 2 store managers in this environment can indeed be assumed to have considerable career concerns. In fact, the district manager's assessment of the store managers' performance is of substantial importance for store managers' career: As Table A7 in the Appendix, for instance, shows, an annual performance evaluation by the district manager is highly predictive for the store manager's future wages. A similar logic applies when the performance reviews themselves generate image concerns as managers do not want to embarrass themselves when not being able to state convincingly that they undertook activities to raise profits.

Hypothesis (iii) reflects the idea that monitoring and performance pay are substitutes as both provide alternative sources of incentives.²⁶ This idea can be formalized in different ways. If, for instance, there is an ideal set of actions that agents can undertake to maximize store profits and moral hazard problems prevent them to do so, then both instruments can move them closer to this level. In Appendix 7.1 we provide a very simple formal model in which we illustrate this mechanism and show the substitution effect. An alternative explanation directly follows from the standard career concerns model: When the review meetings increase the observability of effort, they trigger career concerns to a stronger extent,²⁷ and as shown by Gibbons and Murphy (1992) explicit incentives through performance pay and career concerns are (imperfect) substitutes to raise performance.²⁸

²⁶ Note, that we do not claim that all forms of leadership interactions that affect employee motivation are substitutes to performance pay. Kvaløy et al. (2015) for instance find that motivational talk and performance pay can be complements.

²⁷ In his Nobel lecture Holmström (2017, p 1772) for instance describes the role of career concerns that are guided by the measurement of performance through supervisors such that “[...] a mere change in the accounting system can have a big impact on the behavior of employees”.

²⁸ When performance pay is introduced into this model as in Gibbons and Murphy (1992), efforts are increasing in both the size of a bonus and the precision of the signal. Moreover, whenever the agent's effort cost function has a positive third derivative (which seems a sensible assumption for all tasks where agents cannot increase their efforts indefinitely) the marginal impact of the bonus on efforts will be strictly smaller when reputational incentives are stronger.

4. Results

4.1. Main Treatment Effects

First, it is instructive to consider the number of conversations conducted and the actual bonuses paid out. On average, store managers had 3.5 review conversations with their district managers within the three months period in the respective treatment groups (median=4, SD=1.63) and 91.52% of the first conversations took place within the first two weeks after the start of the project. The average total bonus payment was €535.19 (median=421.65, SD=506.73) with only 17 of the 117 store managers receiving no bonus at all. This bonus payment (€178.40 per month) represents approximately 5.56% of the store managers' gross monthly salary.

Table 2 shows the estimated average treatment effects from fixed effects regressions of store profits on the treatment dummies.²⁹ Column 1 shows results of the (store) fixed models controlling for refurbishments of the stores as well as plan values of profits as predicted by the accounting department in the beginning of the year.³⁰ Column 2 additionally controls for district and store manager fixed effects as the allocation of district manager to stores as well as store manager to stores can change over time. Column 4 and 5 use the same specifications with the log of profits as the dependent variable.³¹ Table A3 in the Appendix provides robustness checks with simple OLS regressions using only the treatment time.³²

²⁹ Note that the treatment here consists of asking district managers to conduct the review conversations. As laid out in the previous paragraph not all district managers implemented all 6 planned meetings. Hence, with respect to the effect of the actually conducted meetings the estimates are intention to treat (ITT) estimates. If we, for instance, apply the logic of a simple Wald estimator our estimates underestimate the effect of fully enforced meetings. However, we caution that an IV strategy instrumenting the conducted meetings with the treatment dummy would not be valid here: as store managers are aware that district managers are supposed to conduct the review meetings, the exclusion restriction would be violated.

³⁰ Some of the stores were refurbished before the intervention. Refurbishment controls include a dummy indicating whether a refurbishment took place in the given month and a dummy indicating whether the store has been refurbished. Not including the controls for refurbishments still leads to qualitatively similar results.

³¹ As ex-post power calculations are criticized (see, e.g., Hoening and Heisey 2001) we further display 90% confidence bands to illustrate the range of possible treatment effects. When using the control group mean over three months prior the experiment to give an intuition about the statistical power, we obtain 72,35% power for an effects size of 0.5 standard deviations.

³² As the 31 clusters in our estimation might seem as a lower bound of the amount of clusters needed, we also estimate standard errors with wild bootstrapping (Wu 1986). The results remain qualitatively robust. For instance, the p -values for the treatment effect in column 2 are: BONUS p -value=0.8438, REVIEW p -value=0.0260, BONUS&REVIEW p -value=0.5085. Using two-way clustered standard errors on districts and time, the results also remain qualitatively robust. For instance, the p -values for the treatment effect in column 2 are: BONUS p -value=0.795, REVIEW p -value=0.023, BONUS&REVIEW p -value=0.455

Table 2: Main Treatment Effects on Profits

	(1) Profits	(2) Profits	(3) CI 90%	(4) Log (Profits)	(5) Log (Profits)	(6) CI 90%
Treatment Effect BONUS	-51.85 (607.3)	156.2 (710.5)	[-1049.6; 1362.7]	-0.00441 (0.0417)	0.0141 (0.0569)	[-0.0825; 0.1108]
Treatment Effect REVIEW	1370.2** (559.0)	1492.3** (666.2)	[361.6; 2622.9]	0.0732*** (0.0238)	0.0858** (0.0411)	[0.0161; 0.1554]
Treatment Effect BONUS&REVIEW	-376.3 (605.1)	-397.7 (564.3)	[-1355.5; 560.0]	-0.00485 (0.0351)	-0.00390 (0.0501)	[-0.0889; 0.0811]
Wald test REVIEW=BONUS&REVIEW	$p=0.0162$	$p=0.0090$		$p=0.0218$	$p=0.0330$	
Time FE	Yes	Yes		Yes	Yes	
Store FE	Yes	Yes		Yes	Yes	
District Manager FE	No	Yes		No	Yes	
Store Manager FE	No	Yes		No	Yes	
Refurbishments	Yes	Yes		Yes	Yes	
Planned Profits	Yes	Yes		Yes	Yes	
N of Observations	3975	3777		3966	3768	
N of Stores	224	224		224	224	
Cluster	31	31		31	31	
Within R^2	0.2370	0.2722		0.1621	0.1875	
Overall R^2	0.7577	0.5955		0.6158	0.4316	

Note: The table reports results from a fixed effects regression with the profits on the store level as the dependent variable. The regression accounts for time and store fixed effects and adds fixed effects for district manager and store managers in column 2&4. The regressions compare pre-treatment observations (January 2016 - March 2017) with the observations during the experiment (April 2017 – June 2017). *Treatment Effect* thus refers to the difference-in-difference estimator. All regressions control for possible refurbishments of a store and the companies planned value of profits. Observations are excluded when a store manager switched the store during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. Columns (3) and (6) display 90% confidence intervals of the specification in columns (2) and (5), respectively. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 2 shows our three main results: Bonuses do not raise performance. Performance reviews increase performance quite substantially, but the latter positive effect of performance reviews is destroyed when store managers also receive performance bonuses.

A first insight is thus that the *BONUS* treatment does not have a significant effect on store profits relative to the control group in all specifications. To understand this, it is first important to recall that also managers in the control group received an information package including an online training of the new profit metric and information about profit margins of the different products. The absence of a significant bonus effect thus shows that the bonus did not raise performance above any effect of this information. In a companion paper (Manthei et al. 2021b) we evaluate the impact of the information package in a field experiment in another region of the same company. In that experiment we found that when the bonus and the information package were introduced separately both raised performance above a control group of stores receiving neither. But the combination of both did not yield significant performance

increases above the effects of the separate practices. The key argument put forward in that paper is that the bonus and the provided information are to some extent substitutes as they both generate attention for the objective to raise profits.

But as column 1 shows, performance reviews are powerful in generating returns above the effect of the information intervention. *REVIEW* significantly increases monthly profits relative to the control group by, on average, €1370. The result remains robust when including store manager and district manager fixed effects. Column 2 displays an estimated treatment effect of €1,492. According to the log specification in columns (4) and (5) performance reviews increase profits by about 7%. Of course, the performance reviews also come at a cost which is essentially the time invested by district and store managers in the conversations and their opportunity costs. Approximating the duration of the conversation with a maximum of 30 minutes for each and using hourly wages of store and district managers, the opportunity cost of a meeting is less than €40. Hence, the total time costs for the pure *REVIEW* intervention are substantially smaller than the estimated treatment effects.³³ Hence, the introduction of performance review conversations about a specific key figure increased this key figure considerably at rather low costs. That is, simple conversations in which store managers have to report their activities to raise profits to their supervisors have a substantial effect on performance.

Our third main result is that the *BONUS&REVIEW* treatment has no significant effect on the outcome variable relative to the control group in any specification. Importantly, in all specifications the treatment effect of *BONUS&REVIEW* is significantly smaller than that of *REVIEW* (Wald test, $p < 0.05$). We thus find – in contrast to our initial hypothesis – that the use of performance pay did not only reduce the *marginal* impact of performance reviews but even lowered their absolute effect.³⁴ In other words: the bonus eliminated the value of the performance review conversations entirely. We explore reasons for this finding in the following.

³³ Recall that district managers visit the stores anyway during the week so that typically no additional travel costs occurred. However, we acknowledge that there may be other unobserved opportunity costs when the time invested in the review conversations is subtracted from the time budget for other activities. But as long as such effects are accounted for in changes in store profits during the time of the intervention they are incorporated in the above calculation.

³⁴ To exclude that the results are driven by outliers in specific districts or specific characteristics of district managers we also ran this regression repeatedly excluding each single district in one regression and the result remains stable.

4.2. How did Bonus Payments Undermine the Benefits of Performance Reviews?

The key question is now how the use of performance pay undermined the benefits of performance reviews. A number of papers (e.g. Gneezy and Rustichini 2000a, 2000b, Bénabou and Tirole 2003, 2006, Fehr and Rockenbach 2003, Fehr and List 2004, Falk and Kosfeld 2006, Sliwka 2007, Ellingsen and Johannesson 2008, Ariely et al. 2009, Christ 2013, Cardinaels and Yin 2015, Alfitian et al. 2021) have for instance argued that bonus payments can sometimes weaken intrinsic incentives to perform a task. Yet, this raises the question why such crowding-out effects should be particularly strong when performance review conversations are used and, moreover, which specific activities are undermined.

We furthermore acknowledge that in contrast to the bonus which is very precisely defined and implemented in exactly the same manner for all store managers, review conversations are by design conducted by supervisors who are human beings and thus may vary in their styles of execution. In the following, we investigate the effects of the bonus payment on the nature of the conversations and possible behavioral channels for the undermining effect. We first study the timing of the effect and then collect evidence from surveys to build a better understanding of what managers actually undertook to raise profits. We then analyze the quantity and content of the review conversations in order to investigate the impact of the bonus on the nature of the review conversations. Moreover, we analyze post-experimental questionnaire data on the managers' satisfaction, their relationship to their district managers, and their perception of the project evaluated in this study.³⁵

4.2.1. The Timing of the Effect

A first key question is whether the review meetings affected behavior through a change in incentives – as we hypothesized at the outset – or a change in the district managers' human capital investments in the store managers' ability to raise profits. In order to explore this, it is instructive to consider the performance effects of the treatment over time including the months after the end of the treatment. Recall that the intervention lasted three months, i.e. store managers received the bonus during this period and review meetings were supposed to take place biweekly during these months in the respective treatments. If the review meetings raised the store managers' human capital, we should see persistent profit increases after the end of the

³⁵ Of course, we acknowledge that this section is exploratory in nature.

treatment period. In case the review meetings rather triggered higher-powered incentives to perform due to reputational concerns and helped store managers to guide these efforts towards the most profitable tasks, treatment differences should vanish after the end of the treatment period.

Table A4 in the Appendix shows treatment effects by the respective months of the intervention including two months after the end of the treatment. The most interesting insight here is that performance effects in the *REVIEW* treatment are strongest in the beginning of the treatment time and there are no post-treatment performance differences. Hence, the meetings clearly have no persistent human capital effects.³⁶ The meetings thus rather affected incentives to exert effort towards tasks that increase profits when store managers knew that they had to report their actions to their supervisors but not beyond. Moreover, the effects are more pronounced in the first two months which further lends support to the role of reputational concerns. In analogy to Holmström's (1999) career concerns model, if agents intend to signal their ability (or more broadly their willingness) to perform activities to raise the key figure, these incentives should be strongest in the very beginning.³⁷

4.2.2. What did the Store Managers do?

A second question is what the store managers actually did to raise profits and whether the treatment differences result from managers focusing on different tasks when they receive a bonus. To investigate the executed tasks in more detail, we first analyze which actions store managers actually undertook in their daily jobs by their own report and what they did to increase profits after the beginning of the experiment. We clustered 29 tasks distilled from job description documents into 7 task dimensions: *personnel management*, *ordering*, *cleanliness*, *inventory management*, *placements*, *analysis of key performance indicators*, and *own operational activities* (cash desk, own customer interaction).³⁸

To assess the relative importance of these tasks, we included 29 items in our post-experimental survey listing the different tasks and asked store managers to state to what extent they generally focused on the respective task in their work in the period of the experiment (1=low focus, 6=high focus). Recall that in all four treatment groups (including the control

³⁶ Note that this also rules out that district managers invested less in the human capital of store managers in the meetings when these store managers received a bonus.

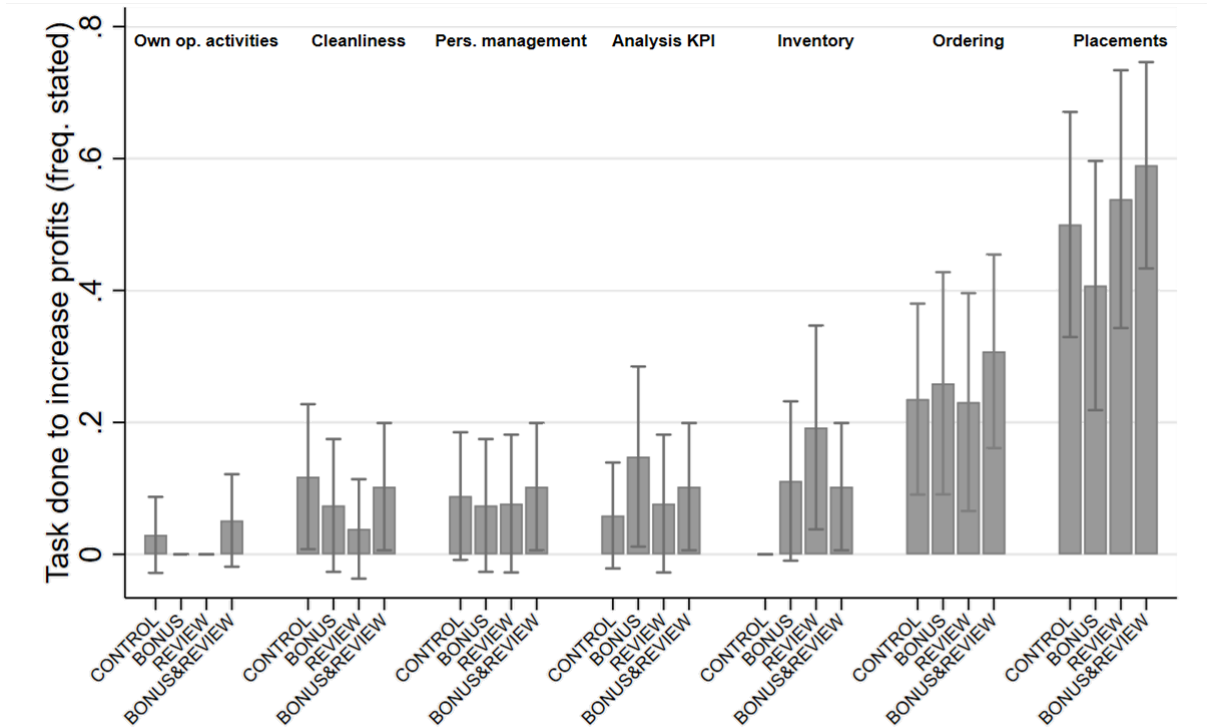
³⁷ The declining performance effects may also to some extent be attributable to fading attention (see Manthei et al. 2021b). Note, however, that in the current experiment, attention for the profit metric is generated for all managers including those in the control group who also received the information package.

³⁸ This classification is displayed in Table A1 in the Appendix.

group) store managers had received an online training on the new profit metric and information on product margins at the beginning of that period. The analysis³⁹ shows that store managers in all groups put a particular focus on ordering (e.g. ordering of meat, vegetables, fruit and bakery products) and inventory management (e.g. analysis of shrinkage, checking of incoming goods).

While the previous items asked about the general focus in the store manager’s work, our survey also included an open-ended question in the survey asking store managers explicitly what they did to increase profits. Research assistants assigned the responses to the 29 task categories. Figure 1 shows the respective frequencies for the different task dimensions.

Figure 1: Self-Reported Tasks Done to Increase Profits



Note: The figure displays the share of stated tasks dimensions to increase profits obtained from open questions of an ex-post questionnaire. 95% confidence bars are displayed.

Most notably, store managers tried to increase profits through improved placements (many for instance stated that they explicitly tried to place articles with high margins in prominent positions), and ordering. Yet, we do not see sizeable systematic treatment differences

³⁹ Figure A1 in the Appendix displays average ratings for the respective task dimensions (normalized by dividing the focus rating for a task by the average focus rating across all tasks)

that stand out and between-task differences in the frequency of the tasks mentioned are much more substantial than within-task treatment differences.⁴⁰

However, we find some indication for slight differences in the relative task focus of store managers when considering the effect of the treatments on the different profit components (see Table A5 in the appendix). The results indicate that store managers in the *REVIEW*-treatment seem to have more aptly managed the balance between reduced personnel costs and an increase in the gross product margin (*Net Sales – Cost of Goods Sold*) selling more profitable products. Store managers in the *BONUS&REVIEW* treatment even reduced personnel costs significantly but failed to raise the gross profit margin and thus did not achieve higher profits which may indicate that the bonus here rather induced “naïve” cost-cutting behavior. Yet we do not observe a similarly pronounced pattern of a reduction in personnel costs at the expense of gross profit margin in the *BONUS* treatment. Hence, it seems unlikely that the difference between *REVIEW* and *BONUS&REVIEW* is entirely driven by naïve cost-cutting induced by bonuses.

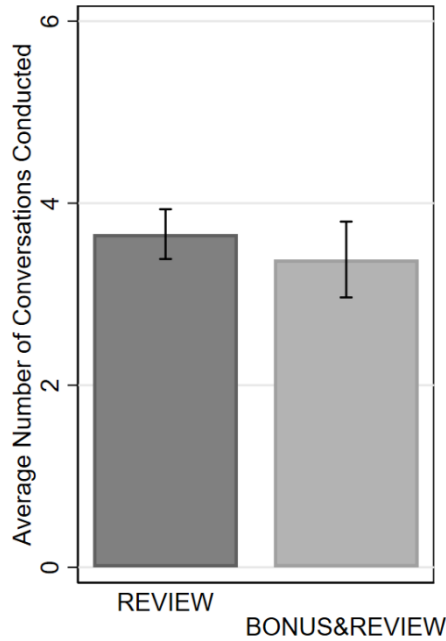
4.2.3. *Frequency of Review Conversations*

Another possible driver of our treatment difference is a potential change in the nature of the review conversations. At our request, the company had asked district managers conducting the review meetings to fill out a short form documenting the contents of the conversation in a concise manner after each meeting. We now use these protocols to assess differences in the way in which the reviews were conducted.

As a first step, we counted the number of meetings and measured the lengths of the protocols (number of notes). As district managers were asked to initiate bi-weekly meetings with their store managers, the maximal number of meetings that could have been conducted was 6 for each store. As stated already in the above, on average across both treatments 3.5 of these 6 meetings were actually conducted by the district managers. As displayed in Figure 2, district managers in the *REVIEW* treatment conducted slightly more conversations (3.66) than in *BONUS&REVIEW* (3.38) but this difference is not significant (MWU, $p=0.4282$; OLS regression with clustered standard errors, $p=0.705$).

⁴⁰ In fact, the only task dimension for which we see a statistically significant difference in the responses between store managers in *REVIEW* as compared to *BONUS&REVIEW* is own operational activities, which is significantly more often mentioned in *BONUS&REVIEW* (as it is never mentioned in *REVIEW*, $p=0.042$ in an OLS regression with clustered standard errors).

Figure 2: Average Number of Conversations Conducted over the three-month Experimental Period



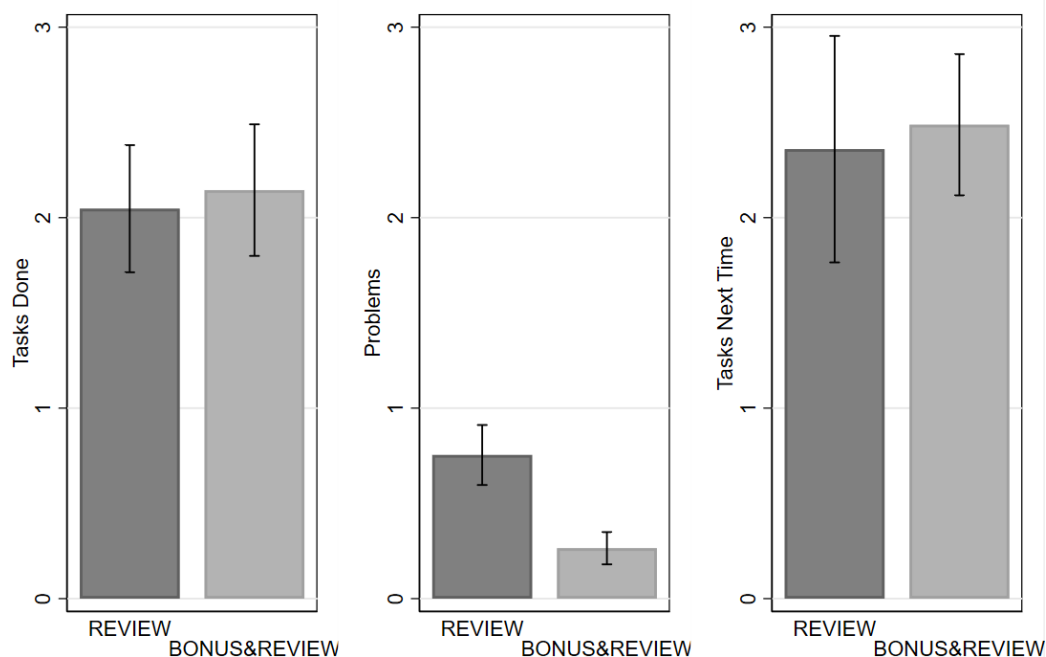
Note: The figure displays the average number of conversations per store manager. 95% confidence bars are displayed.

4.2.4. Content of the Conversations

To explore whether the quality of the review conversations changed we now explore the content of the meetings in more detail. Recall that district managers were asked to go through three sections in the meetings asking store managers about their respective activities. For each of these sections they were asked to protocol what the store managers reported. Figure 3 displays information about the intensity of the protocolled parts of the conversations. It shows the average number of notes in the sections (i) “What did the store manager do to increase profits?” (ii) “Which problems occurred?”, and (iii) “What would the store manager like to do before the next meeting?”⁴¹

⁴¹ “Notes” refers here to bullet points in the protocol. This can be a couple of words or a sentence. Nearly always it refers to one task.

Figure 3: Average Number of Notes per Conversation over the three-month Experimental Period



Note: The figure displays the per session average number of notes/sentences per store manager in the respective category. 95% confidence bars are displayed.

A first key observation is that there are no significant differences in sections (i) on the tasks done and (iii) regarding the tasks planned – if anything there are even more activities reported when there is a bonus. However, there is a sizeable difference for category (ii): The use of performance pay substantially reduced the number of problems stated by store managers during the conversations as shown in the second panel of Figure 5. In *BONUS&REVIEW* store managers only state on average 0.27 problems per conversations, in *REVIEW* store managers state nearly three times as many (0.75) problems per conversation (MWU, p -value<0.001).⁴² In fact, for 60.3% of the stores not a single problem was mentioned in any of the meetings in *BONUS&REVIEW* while this fraction is only 22.2% in the *REVIEW* treatment.⁴³ Table 3 shows treatment effect from OLS Regressions.

⁴² As Figure A3 in the appendix shows, the effect is not driven by single district managers, but the pattern is very similar across district managers of the respective groups (MWU test with one observation per district manager averaged across all stores, $p=0.0421$). Figure A4 in the appendix shows the timing of stated problems within conversations.

⁴³ We also explored the specific content of the meetings. Student helpers classified the notes again into the 29 different tasks for each of the up to 6 meetings held in each store. Figure A5 in the Appendix displays the respective frequencies of mentioning a topic in the seven tasks dimensions (summed up across all three sections of the protocol). The ranking of the task dimensions is well in line with the ranking of the relative importance from the open-ended question to store managers. When bonuses are

Table 3: Treatment Effects on Review Conversation Notes

<i>Reference Group:</i> <i>Treatment REVIEW</i>	(1) Overall	(3) Tasks Done	(5) Problems	(7) Tasks Next Time
Treatment Effect BONUS&REVIEW	0.166 (0.765)	0.0854 (0.734)	-0.482*** (0.140)	0.136 (0.628)
N of Observations	118	118	118	118
Cluster	18	18	18	18
<i>Overall R</i> ²	0.0018	0.008	0.1623	0.0010

Note: The table reports results from ordinary least squares (OLS) regressions with the different subsections of the review conversations as depending variable. Regressions further control for store size, number of employees, store manager’s age and prior performance evaluation, as well as randomization group. The Treatment REVIEW serves as the reference group. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Hence, the bonus seems to have shifted the incentives to talk openly about encountered problems. Note that this effect is also still significant when we account for multiple hypotheses testing here. A simple Bonferoni correction (Bonferoni 1935) considering the three tests we run still results in a p-value of < 0.001 . The question remains whether the substantially lower number of mentioned problems in the review meetings is indeed associated to the performance loss when bonuses are in place. To explore this, we now consider only the subset of stores from the *REVIEW* and *BONUS&REVIEW* treatments in which at least one problem has been mentioned in a review conversation. In other words, we include only review meetings that entailed sufficiently “open” conversations. We then replicate the results from Table 2 on the reduced data set and report the results in Table 4.

in place, review meetings tend to be more concerned with placements, personnel management and own operational activities and less concerned with ordering behavior and cleanliness (see Table A6 in the Appendix for regression results).

Table 4: Main Treatment Effects on Profits (only reviews where problems mentioned)

	(1)	(2)	(3)	(4)
	Profits	Profits	Log (Profits)	Log (Profits)
Treatment Effect BONUS	-0.0126 (0.620)	0.207 (0.713)	-0.00484 (0.0421)	0.0147 (0.0572)
Treatment Effect REVIEW	1.261** (0.531)	1.402** (0.596)	0.0719*** (0.0213)	0.0827** (0.0373)
Treatment Effect BONUS&REVIEW	0.643 (0.706)	1.126 (0.745)	0.0515* (0.0268)	0.0687 (0.0415)
Wald test REVIEW=BONUS&REVIEW	$p=0.4070$	$p=0.7168$	$p=0.3436$	$p=0.5929$
Time FE	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes
District Manager FE	No	Yes	No	Yes
Store Manager FE	No	Yes	No	Yes
Refurbishments	Yes	Yes	Yes	Yes
Planned Profits	Yes	Yes	Yes	Yes
N of Observations	3046	2917	3040	2911
N of Stores	172	172	172	172
Cluster	30	30	30	30
Within R^2	0.301	0.354	0.158	0.183
Overall R^2	0.822	0.748	0.606	0.522

Note: The table reports results from a fixed effects regression with the profits on the store level as the dependent variable. The regression accounts for time and store fixed effects and adds fixed effects for district manager and store managers in column 2&4. The regressions compare pre-treatment observations (January 2016 - March 2017) with the observations during the experiment (April 2017 – June 2017). *Treatment Effect* thus refers to the difference-in-difference estimator. All regressions control for possible refurbishments of a store and the companies planned value of profits. Observations are excluded when a store manager switched the store during the treatment period. Observations are further excluded if no problem was mentioned in any performance review. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

The coefficients of *BONUS&REVIEW* are now considerably larger than in the full sample and the treatment effect of 5.2% (as for instance estimated in the log specification in column (3)) now moves much closer to those of the *REVIEW* treatment. This indicates mentioning problems openly is indeed an indicator for the quality of review conversations and bonus payments undermined this quality. Importantly, in none of the specifications the coefficient of *BONUS&REVIEW* is now significantly smaller than that of *REVIEW*.

4.2.5. *Post-Experimental Questionnaire: Satisfaction and Feedback Quality*

In an additional part of the survey we asked store managers about their overall satisfaction with their job as well as specific job domains such as their compensation and their workload. As columns (1)-(3) in Table 5, which display results from regressions of the

respective survey items on treatment dummies, show, the treatments did not affect employee satisfaction in a detectable manner.

The survey then includes an item about their own perceived aim to raise profits (“*I have tried to increase profits in the last few months*”) as well as items eliciting store managers’ perceptions on the interaction with their respective district manager (“*My district manager gave me regular feedback*”, “*My district manager motivated me regularly to do better*”). Columns (4)-(6) of Table 5 report the respective regression results.

Table 5: Survey Results Perceptions on Activities

	(1) Satisfaction Job	(2) Satisfaction Compens.	(3) Satisfaction Workload	(4) Profit Aim	(5) Feedback	(6) Motivate
Treatment Effect BONUS	-0.313 (0.304)	0.341 (0.262)	-0.257 (0.550)	-0.0657 (0.268)	0.138 (0.269)	0.289 (0.345)
Treatment Effect REVIEW	0.114 (0.254)	-0.031 (0.358)	-0.314 (0.554)	0.128 (0.249)	0.931*** (0.304)	0.831* (0.445)
Treatment Effect BONUS&REVIEW	-0.133 (0.311)	0.0138 (0.228)	-0.551 (0.445)	0.538** (0.236)	0.385 (0.248)	0.00922 (0.343)
Wald test REVIEW=BONUS&REV.	$p=0.3986$	$p=0.8949$	$p=0.5904$	$p=0.1304$	$p=0.0428$	$p=0.0372$
N of Observations	97	97	97	95	96	96
Cluster	28	28	28	28	28	28
Overall R^2	0.140	0.303	0.093	0.177	0.189	0.174

Note: The table reports results from OLS regressions with the respective survey response as the dependent variable (scale from 1-6). “Job” is general job satisfaction, “Compens.” is satisfaction with the compensation and “Workload” is satisfaction with the workload. Further controls are store size, number of employees, store manager’s age and prior performance evaluation, as well as randomization group. Standard errors are clustered on the district level of the treatment start and displayed in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

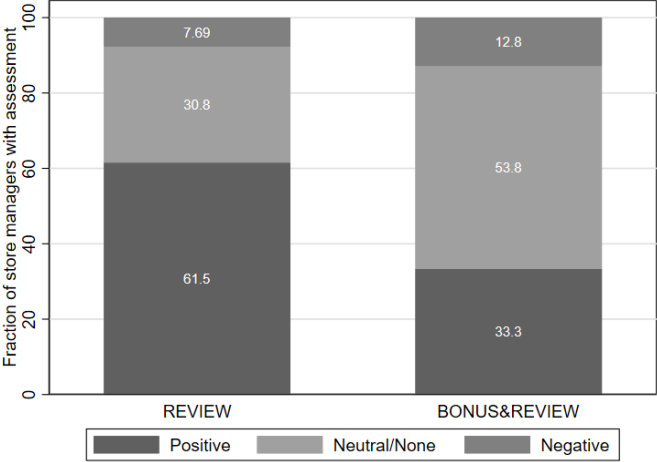
In contrast to the results on satisfaction, here we observe specific patterns in the store managers’ perceptions. Store managers state the strongest aim to increase profits when they receive a bonus and have review meetings (column (4) of Table 5).⁴⁴ As columns (5) and (6) show, however, the store manager’s perception that they receive regular feedback from and feel motivated by their district managers is highest in the *REVIEW* treatment. And these positive effects vanish when they also receive a bonus: The coefficient of the *REVIEW* treatment is significantly larger than that of the *BONUS&REVIEW* treatment ($p=0.0428$ and $p=0.0372$ for the feedback and motivation item respectively). This lends further support to the idea that bonus

⁴⁴ It is conceivable that this due to demand/social desirability effects: this group had the most intensive set of practices to increase profits (and the costliest investment by the firm): hence, store managers may have felt the obligation to state the strongest consent when they receive both a bonus and the review meetings.

payments changed the nature of the feedback conversations and reduced the quality of interaction.

Moreover, we asked store managers an open question about their opinion on the project (“How did you perceive the regular conversations with your district manager?”) and categorized the answers into positive, neutral/none and negative. The results are displayed in Figure 4. Of all store managers responding to the survey, 61.5% stated a positive opinion in the *REVIEW* treatment against only 33.3% in the *BONUS&REVIEW* treatment (MWU, $p=0.0262$; OLS Regression with clustered standard errors, $p=0.025$). The fraction of negative assessments is higher in the *BONUS&REVIEW* treatment, but this difference is not statistically significant (MWU, $p= 0.5168$; OLS Regression with clustered standard errors, $p= 0.420$).

Figure 4: Project Assessment by Store Managers



Note: The figure displays the fraction of store managers with a positive, neutral/none, or negative assessment about the project stated in an open question in the post-experimental survey. N=65.

4.2.6. Discussion of the Additional Findings

Comparing the review treatments with and without the bonus we thus have found the following pieces of evidence:

- The number of conducted meetings is only slightly (and statistically insignificantly) smaller when bonuses are paid.

- There are no distinctive differences in the tasks on which store managers focus in the intervention period.
- But when store managers receive no bonus, they
 - mention encountered problems substantially more often in the reviews (by a factor of three) while reporting a similar number of undertaken and planned activities,
 - perceive to have received better feedback and feel more motivated by their district managers,
 - and liked the project more even though they earn less.
- Moreover: The negative effect of the bonus vanishes when we include only observations from *BONUS&REVIEW* in which at least one problem is mentioned.

The bonus thus did not affect the number of reported and planned activities (if anything, store managers reported slightly more activities in *BONUS&REVIEW*) but it reduced the number of mentioned problems substantially. A key difference between reporting an activity and reporting a problem to a supervisor is that the latter leads to a stronger involvement of this supervisor (or may even create an obligation for this supervisor to act). It therefore imposes a stronger externality on the respective supervisor. The bonus payments thus have shifted the nature of the review conversations towards a stronger self-reliance of store managers.

An alternative explanation could be that the behavioral shift is driven by a reduction in effort of district managers. For instance, when store managers receive bonuses district managers may have felt less responsible for raising profits, reducing their own input. Or supervisors may have reduced their efforts out of envy towards store managers because of the bonus. However, this seems unlikely for several reasons. For one, we do not see that they initiated a significantly smaller number of review meetings. Moreover, as stated above, district managers' salaries are nearly twice as large as those of store managers. District managers also receive a bonus which depends on the profit of the region and their own performance is evaluated based on the success of the stores they manage. Hence, by reducing efforts in this respect, district managers would have harmed themselves. Finally, if district managers indeed actively provided less input when store managers receive a bonus we should see in particular a lower number of planned activities

in the review meetings as here district managers could have provided the most input suggesting further activities.

The shift towards a stronger self-reliance of store managers is in line with a set of experimental results in psychology showing that priming subjects with the concept of money creates a more self-sufficient orientation (Vohs et al. 2006, Vohs 2015).⁴⁵ They find that reminders of money led subjects in different experimental settings to pose fewer requests for help to solve a problem and to prefer working independently. A related economic reasoning follows from the interplay of monetary and career concerns/image motivation. Bénabou and Tirole (2006) analyze a model in which agents have reputational concerns to be viewed as being motivated showing that in such context bonuses can naturally change the reputational incentives to perform an activity as they may “*create doubt about the true motive*” (Bénabou and Tirole 2006, p. 1652) for which an action is taken.

In Appendix 7.2 we develop a formal model applying this idea to study the differential effect of bonus payments on the incentives to perform tasks that either impose an externality on the supervisor or are performed independently by the agent. The idea is simple: in the model the agent works on two tasks. On one task the agent can work by himself, but the other one involves the supervisor: higher efforts by the agent here create an externality on the supervisor (such as when stating a problem or asking for help). The agent has image concerns and is interested in showing the supervisor that he wants “to do a good job” – which here means choosing actions that tend to maximize overall surplus. We then show that image motivation – that may be triggered by performance reviews – can raise incentives for both tasks if there is no bonus. When, however, a bonus is used the agent (rationally) fears that higher efforts for the task that imposes the externality may be viewed as being selfishly motivated. The analysis now shows that this naturally drives a wedge between the signaling values of both actions. In equilibrium, image concerns still raise incentives for the action that can be performed independently but lower incentives for the task that involves the supervisor. Moreover, the bonus eliminates the overall performance gain induced by image motivation.

This crowding-out effect should therefore be pronounced when (i) there are large returns to reputation, (ii) supervisors are able to observe efforts precisely and (iii) for tasks which impose an externality on others. In order to assess the first condition, recall the considerable career incentives in our setting as outlined in section 2: gains from promotion are substantial

⁴⁵ See also Genschow et al. (2020) for a recent discussion of the robustness of these findings.

(raising wages by up to 100%) and supervisor evaluations are strongly predictive for future wage increases (see Table A7). As to the second condition: By generating transparency about efforts performance reviews most likely augment reputational incentives. By the same token, detrimental effects of performance pay on reputational incentives are then naturally stronger when performance reviews are in place.

Finally, the model rationalizes why the crowding-out effect becomes salient in particular through a reduction of open statements about encountered problems. In this respect, consider a store manager who raises a problem in a review conversation. Without bonus payments such a statement of problems should be a rather genuine signal of the motivation (or ability) to do a good job. If, however, performance pay is in place, the store manager may rationally fear that stating a problem is perceived to be driven by mere self-interest. Another way of putting this is that the bonus changed the social norms of behavior⁴⁶ in the review meetings towards a stronger self-reliance. Note that materially store managers should have benefited from open discussions about encountered problems in order to lead district managers to help them and guide their efforts towards profitable tasks – thus raising profits and, in turn, their own bonuses. However, store managers did the opposite, reporting a lower number of problems. The stronger self-reliance induced by the bonus thus apparently led to less open discussions and a lower feedback quality – and this seems to have backfired.

5. Conclusion

Implementing appropriate management practices to align the behavior of employees with the interests of the employer is one of the biggest challenges in the design of organizations. We have conducted a field experiment investigating the impact of two such practices, bonus payments and performance reviews, as well as their interplay.

Our paper provides three key results. First, regular conversations between supervisors and subordinates about activities to raise profits can lead to substantial profit increases. In other words, forcing managers to explain what they do to raise a specific outcome variable can be very effective in increasing this variable. Second, this effect surpasses the effects of substantial bonus payments for the same key figure. Thus, we may conclude that indeed talking about

⁴⁶ See Bénabou and Tirole (2011) or Ali and Bénabou (2020) for models that formalize a notion of social norms as determined by reputational concerns.

performance is more worthwhile than merely paying for it. These results are in line with the view put forward in the career concerns literature that reputational incentives can be a powerful motivational force in organizations. Moreover, by designing evaluation practices organizations can shift reputational incentives towards specific performance objectives – without the necessity to spend money on this.

Our third result is that the use of performance pay substantially reduced the value of the review conversations. Investigating the underlying behavioral channels exploring data from surveys and conversation protocols, we found evidence that the use of bonus payments led to a stronger self-reliance of store managers who raised a substantially lower number of problems in the review meetings. We moreover provide a rationale for this finding in a formal multitasking model where we show that – when agents have image concerns – bonuses can indeed naturally generate a stronger self-reliance as agent rationally fear that activities that involve others may be interpreted as being driven by mere self-interest. This, stronger self-reliance in turn can reduce positive reputational incentives triggered by performance review meetings.

Given that performance review conversations alone were so effective in raising profits, this certainly raises the question why the firm had not implemented such meetings beforehand. Here it is important to stress that, as laid out in the above, structured meetings between store managers and their supervisors had been conducted regularly also prior the experiment. However, our intervention explicitly introduced meetings that were intentionally designed to talk about activities to raise profits as measured by a specific metric. What we thus show is that conversations *about a specific performance objective* can be very effective in yielding this objective.

Our results also shed some light on a recent debate on the use of performance feedback and performance pay in firms. Traditionally, performance reviews have often been used to assess performance and allocate bonuses (see, e.g. Cappelli and Conyon 2018). In recent years, however, many firms have intentionally shifted the focus in performance reviews away from the allocation of rewards. Several larger companies have entirely stopped or strongly revised standard annual performance ratings and instead established regular feedback conversations (see, for instance, Buckingham and Goodall 2015, Cappelli and Tavis 2016). Frequently, this change has been triggered by a feeling that a continuous dialogue between supervisor and subordinate is a key driver for performance and may be more important than incentives set

through evaluation and compensation. Our evidence shows that monetary rewards based on individual performance may even undermine the objective to foster an open dialogue.

On a more general level, our results show that different organizational practices may interact in non-trivial ways. As has been stressed in the literature on complementarity in organizations⁴⁷, the performance effect of introducing a specific management practice may be contingent on the use of other practices. Whether and how specific practices interact depends on the interplay of different economic motives and behavioral mechanisms. Brynjolfsson and Milgrom (2013) describe challenges in the empirical assessment of interdependencies between organizational practices, stating that the opportunities to run designed experiments in firms are “underexploited” in this respect. RCTs that simultaneously vary the use of two practices are still rare, but can advance our understanding of the role of such interdependencies for firm performance and at the same time allow studying the relevance of different behavioral mechanisms in field settings.

⁴⁷ See, e.g. Milgrom and Roberts (1990, 1995), Ichniowski et al. (1997), Ferreira and Otley (2009), Grabner and Moers (2013), or Brynjolfsson and Milgrom (2013) for a recent survey.

6. References

- Alftian, Jakob, Dirk Sliwka, and Timo Vogelsang. "When Bonuses Backfire: Evidence from the Workplace." *Working Paper* (2021).
- Ali, S. Nageeb, and Roland Bénabou. "Image versus information: Changing societal norms and optimal privacy." *American Economic Journal: Microeconomics*, 12.3 (2016): 116-64.
- Antonakis, John, Giovanna d'Adda, Roberto Weber, and Christian Zehnder. "Just words? Just speeches? On the economic value of charismatic leadership." *Working Paper* (2021).
- Ariely, Dan, Anat Bracha, and Stephan Meier. "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *American Economic Review* 99.1 (2009): 544-55.
- Ashraf, Nava, Oriana Bandiera, and Kelsey Jack. "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of Public Economics* 120 (2014): 1-17.
- Athey, Susan and Guido Imbens. "The Econometrics of Randomized Experiments.", in A. Banerjee and E. Duflo, *Handbook of Field Experiments* 1 (2017): 73-140.
- 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, Andrea Prat, Stephen Hansen, and Raffaella Sadun. "CEO behavior and firm performance." *Journal of Political Economy* 128.4 (2020): 1325-1369.
- Banker, Rajiv D., Seok-Young Lee, Gordon Potter, and Dhinu Srinivasan. "An Empirical Analysis of Continuing Improvements Following the Implementation of a Performance-Based Compensation Plan." *Journal of Accounting and Economics* 30.3 (2000): 315-350.
- Banker, Rajiv D., Shunlan Fang, and Seok-Young Lee. "Conflict between Supervisory Monitoring and Monetary Incentives – Evidence from a High-End Retail Store." *Working Paper* (2018).
- Barankay, Iwan. "Rank Incentives: Evidence from a Randomized Workplace Experiment." *Working Paper* (2012).
- Bartel, Ann, Casey Ichniowski, and Kathryn Shaw. "Using 'insider econometrics' to study productivity." *American Economic Review Papers and Proceedings* 94.2 (2004): 217-223.
- Belot, Michèle, and Marina Schröder. "The spillover effects of monitoring: A field experiment." *Management Science* 62.1 (2016): 37-45.
- Bénabou, Roland, and Jean Tirole. "Intrinsic and extrinsic motivation." *The Review of Economic Studies* 70.3 (2003): 489-520.

- Bénabou, Roland and Jean Tirole. "Incentives and Prosocial Behavior." *American Economic Review* 96.5. (2006): 1652-1678.
- Bénabou, Roland and Jean Tirole. "Laws and Norms" IZA Discussion Paper 6290 (2012).
- Bender, Stefan, Nicholas Bloom, David Card, John Van Reenen, and Stefanie Wolter. "Management practices, workforce selection, and productivity." *Journal of Labor Economics* 36.1 (2018): 371-409.
- Bertrand, Marianne, and Antoinette Schoar. "Managing with style: The effect of managers on firm policies." *Quarterly Journal of Economics* 118.4 (2003): 1169-1208.
- Buckingham, Marcus, and Ashley Goodall. "Reinventing performance management." *Harvard Business Review* 93, no. 4 (2015): 40-50.
- Bursztyjn, Leonardo, and Robert Jensen. "How Does Peer Pressure Affect Educational Investments?" *Quarterly Journal of Economics*, 130. 3 (2015): 1329–1367.
- Blader, Steven, Claudine Gartenberg, and Andrea Prat. "The contingent effect of management practices." *The Review of Economic Studies* 87.2 (2020): 721-749.
- Blanes i Vidal, Jordi and Mareike Nossol. "Tournament Without Prizes: Evidence from Personnel Records." *Management Science* 57.10 (2011): 1721-1736.
- Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128.1 (2013): 1-51.
- Bloom, Nicholas, James Liang, John Roberts, and Zhichung Jenny Ying. "Does Working from Home Work? Evidence from a Chinese Experiment." *Quarterly Journal of Economics* 130.1 (2015): 165-217.
- Bloom, Nicholas, and John Van Reenen. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122.4 (2007): 1351-1408.
- Boly, Amadou. "On the incentive effects of monitoring: evidence from the lab and the field." *Experimental Economics* 14.2 (2011): 241-253.
- Bonferroni, Carlo E. "Il calcolo delle assicurazioni su gruppi di teste." *In Studi in Onore del Professore Salvatore Ortu Carboni* (1935): 13–60.
- Bradler, Christiane, Robert Dur, Susanne Neckermann, Arjan Non. "Employee Recognition and Performance: A Field Experiment." *Management Science* 62.11 (2016): 3085-3391.
- Brynjolfsson, Erik, and Paul Milgrom. "Complementarity in Organizations." *Handbook of Organizational Economics* (2013): 11-55.
- Campbell, Dennis, Marc J. Epstein, and F. Asis Martinez-Jerez. "The Learning Effects of Monitoring." *The Accounting Review* 86.6 (2011):1909-1934.
- Cappelli, Peter, and Martin J. Conyon. "What Do Performance Appraisals Do?" *ILR Review* 71.1 (2018): 88-116.
- Cappelli, Peter, and Anna Tavis. "The performance management revolution." *Harvard Business Review* 94.10 (2016): 58-67.

- Cardinaels, Eddy, and Huaxiang Yin. "Think twice before going for incentives: Social norms and the principal's decision on compensation contracts." *Journal of Accounting Research* 53.5 (2015): 985-1015.
- Casas-Arce, Pablo, F. Asis Martínez-Jeres, and V. G. Narayanan. "The Impact of Forward-Looking Metrics on Employees Decision-Making: The Case of Customer Lifetime Value." *The Accounting Review* 92.3 (2017a): 31-56.
- Casas-Arce, Pablo, Sofia M. Lourenço, and F. Asis Martínez-Jeres. "The Performance Effect of Feedback Frequency and Detail: Evidence from a Field Experiment in Customer Satisfaction." *Journal of Accounting Research* 55.5 (2017b): 1051-1088.
- Casas-Arce, Pablo, F. Asis Martínez-Jeres, and Joseph Moran. "Motivating through Managing by Walking Around." *Working Paper* (2019).
- Cassar, Lea, and Stephan Meier. "Nonmonetary Incentives and the Implications of Work as a Source of Meaning." *Journal of Economic Perspectives* 32.3 (2018): 215-38.
- Cai, Jing and Shing-Yi Wang, "Improving Management through worker evaluations: Evidence from Auto Manufacturing" *NBER Working Paper* 27680 (2020).
- Christ, Margaret H. "An experimental investigation of the interactions among intentions, reciprocity, and control." *Journal of Management Accounting Research* 25.1 (2013): 169-197.
- 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.
- Didonato, Tom. "Stop Basing Pay on Performance Reviews." *Harvard Business Review Digital Articles* (2014).
- Eisenhardt, Kathleen M. "Agency theory: An assessment and review". *Academy of Management Review* 14.1 (1989): 57-74.
- Ellingsen, Tore, and Magnus Johannesson. "Paying respect." *Journal of Economic Perspectives* 21.4 (2007): 135-150.
- Ellingsen, Tore, and Magnus Johannesson. "Pride and prejudice: The human side of incentive theory." *American Economic Review* 98.3 (2008): 990-1008.
- Englmaier, Florian, Stefan Grimm, David Schindler, and Simeon Schudy. "The Effect of Incentives in Non-Routine Analytical Team Tasks – Evidence from a Field Experiment." *CESifo Working Paper* No.6903 (2018).
- Eyring, Henry, and V. G. Narayanan. "Performance Effects of Setting a High Reference Point for Peer-Performance Comparison." *Journal of Accounting Research* 56.2 (2018): 581-615.
- Eyring, Henry, Patrick J. Ferguson, and Sebastian Koppers. "Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment." *Journal of Accounting Research* 59.2 (2021): 657-711.

- Fama, Eugene F. "Agency problems and the theory of the firm." *Journal of Political Economy* 88, no. 2 (1980): 288-307.
- Falk, Armin, and Michael Kosfeld. "The Hidden Costs of Control." *American Economic Review* 96.5 (2006):1611-1630.
- Fehr Ernst, and John List. "The Hidden Costs and Returns of Incentives-Trust and Trustworthiness Among CEOs." *Journal of the European Economic Association* 2.5 (2004): 743-771.
- Fehr, Ernst, and Bettina Rockenbach. "Detrimental effects of sanctions on human altruism." *Nature* 422.6928 (2003): 137.
- Ferreira, Aldónio, and David Otley. "The design and use of performance management systems: An extended framework for analysis." *Management accounting research* 20.4 (2009): 263-282.
- Floyd, Eric and John A. List. "Using Field Experiments in Accounting and Finance." *Journal of Accounting Research* 54.2 (2016): 437-475.
- Friebel, Guido, Matthias Heinz, Miriam Krüger, and Nikolay Zubanov. "Team incentives and performance: Evidence from a retail chain." *American Economic Review* 107.8 (2017): 2168-2203.
- Friebel, Guido, Matthias Heinz, and Nikolay Zubanov. "Middle Managers, Personnel Turnover, and Performance: A Long-Term Field Experiment in a Retail Chain." *Management Science* (2021).
- Genschow, Oliver, Johannes Schuler, Emiel Cracco, Marcel Brass, and Michaela Wänke. "The effect of money priming on self-focus in the imitation-inhibition task." *Experimental psychology* (2020).
- Gibbons, Robert, and Kevin J. Murphy. "Optimal incentive contracts in the presence of career concerns: Theory and evidence." *Journal of Political Economy* 100.3 (1992): 468-505.
- Gibbons, Robert and John Roberts. "Economic Theories of Incentives in Organizations," *Handbook of Organizational Economics* (2013): 56-99.
- Gneezy, Uri, and Aldo Rustichini. "Pay enough or don't pay at all." *The Quarterly Journal of Economics* 115.3 (2000a): 791-810.
- Gneezy, Uri, and Aldo Rustichini. "A fine is a price." *The Journal of Legal Studies* 29.1 (2000b): 1-17.
- Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. "When and why incentives (don't) work to modify behavior." *Journal of Economic Perspectives* 25, no. 4 (2011): 191-210.
- Gosnell, Greer K., John A. List, and Robert D. Metcalfe. "The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains." *Journal of Political Economy* 128.4 (2020): 1195-1233.

- Grabner, Isabella, and Frank Moers. "Management control as a system or a package? Conceptual and empirical issues." *Accounting, Organizations and Society* 38.6-7 (2013): 407-419.
- Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. "Learning Through Noticing: Theory and Evidence from a Field Experiment." *The Quarterly Journal of Economics* 129.3 (2014): 1311-1353.
- Harrison, Glenn W. and John List. "Field Experiments." *Journal of Economic Literature* 42.4. (2004): 1009-1055.
- 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.
- Hoffman, Mitchell, and Steven Tadelis. "People management skills, employee attrition, and manager rewards: An empirical analysis." *Journal of Political Economy* 129.1 (2021): 243-285.
- Holmström, Bengt. "Managerial incentive problems: A dynamic perspective." *The Review of Economic Studies* 66.1 (1999): 169-182.
- Holmström, Bengt. "Pay for Performance and Beyond." *American Economic Review* 107(7) (2017): 1753-1777.
- Hossain, Tanjim, and John A. List. "The behaviorist visits the factory: Increasing productivity using simple framing manipulations." *Management Science* 58.12 (2012): 2151-2167.
- Ichino, Andrea, and Gerd Muehlheusser. "How often should you open the door?: Optimal monitoring to screen heterogeneous agents." *Journal of Economic Behavior & Organization* 67.3-4 (2008): 820-831.
- Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi. "The effects of human resource management practices on productivity: A study of steel finishing lines." *American Economic Review* (1997): 291-313.
- Ichniowski, Casey, and Kathryn Shaw. "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices." *Journal of Economic Perspectives* 17.1 (2003):155-180.
- Kvaløy, Ola, Petra Nieken, and Anja Schöttner. "Hidden benefits of reward: A field experiment on motivation and monetary incentives." *European Economic Review* 76 (2015): 188-199.
- Lazear, Edward P. "Performance pay and productivity." *American Economic Review* 90.5 (2000): 1346-1361.
- Lazear, Edward P., and Paul Oyer, "Personnel Economics" in in: R. Gibbons and J. Roberts (eds.), *Handbook of Organizational Economics*. (2013): 479-519.
- Lazear, Edward P., Kathryn L. Shaw, and Christopher T. Stanton. "The value of bosses." *Journal of Labor Economics* 33.4 (2015): 823-861.

- Lazear, Edward P. "Compensation and Incentives in the Workplace" *Journal of Economic Perspectives* 32.3 (2018): 195-214.
- Levitt, Steven D., and Susanne Neckermann. "What field experiments have and have not taught us about managing workers." *Oxford Review of Economic Policy* 30.4 (2015): 639-657.
- List, John A., and Imran Rasul. "Field experiments in labor economics." *Handbook of Labor Economics*. 4. Elsevier (2011): 103-228.
- Lourenço, Sofia M. "Monetary Incentives, Feedback, and Recognition – Complements or Substitutes? Evidence from a Field Experiment in a Retail Service Company." *The Accounting Review* 91.1 (2016): 279-297.
- Manthei, Kathrin, Dirk Sliwka and Timo Vogelsang. "Performance Pay and Prior Learning: Evidence from a Retail Chain." *Management Science* (2021a).
- Manthei, Kathrin, Dirk Sliwka, and Timo Vogelsang. "Information Provision and Incentives - A Field Experiment on Facilitating and Influencing Managers' Decisions." *IZA Discussion Paper* (2021b).
- Milgrom, Paul, and John Roberts. "The economics of modern manufacturing: Technology, strategy, and organization." *American Economic Review* 80.3 (1990): 511-528.
- Milgrom, Paul, and John Roberts. "Complementarities and fit strategy, structure, and organizational change in manufacturing." *Journal of Accounting and Economics* 19.2-3 (1995): 179-208.
- Nagin, Daniel S., James B. Rebitzer, Seth Sanders, and Lowell J. Taylor. "Monitoring, Motivation, and Management: The Determinants of the Opportunistic Behavior in a Field Experiment." *American Economic Review* 92.4 (2002): 850-873.
- Prendergast, Canice. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37.1 (1999): 7-63.
- Rebitzer, James B., and Lowell J. Taylor. "Extrinsic rewards and intrinsic motives: standard and behavioral approaches to agency and labor markets." *Handbook of Labor Economics*, Vol. 4 (2011): 701-772.
- Sliwka, Dirk. "Trust as a signal of a social norm and the hidden costs of incentive schemes." *American Economic Review* 97.3 (2007): 999-1012.
- Tran, Anh, Richard Zeckhauser. "Rank as an inherent incentive: Evidence from a field experiment." *Journal of Public Economics* 96.6 (2012): 645-650.
- Sandvik, Jason, Richard Saouma, Nathan Seegert, Christopher T. Stanton. "Workplace Knowledge Flows." *Quarterly Journal of Economics* 135.3 (2020): 1635–1680.
- Sprinkle, Geoffrey B, and Michael G. Williamson. "Experimental Research in Managerial Accounting." *Handbooks of Management Accounting Research* (2006): 451-444.
- Vohs, Kathleen D., Nicole L. Mead, Miranda R. Goode. "The Psychological Consequences of Money." *Science* 314.5802 (2006): 1154.1156.

Vohs, Kathleen D. "Money priming can change people's thoughts, feelings, motivations, and behaviors: An update on 10 years of experiments." *Journal of Experimental Psychology: General* 144.4 (2015): e86.

Wu, C.F.J. "Jackknife, bootstrap and other resampling methods in regression analysis (with discussions)." *Annals of Statistics* 14 (1986): 1261-1350.

7. Appendix

7.1. A Simple Model of Monitoring and Performance Pay

Consider the following simple extension of a standard linear principal agent model. An agent can exert an effort e to raise store profits π at personal costs $c(e)$ where $c''(e) > 0$ and $c'(\bar{e}) = 0$ for some $\bar{e} > 0$. Profits are given by

$$\pi = e + \varepsilon$$

where ε is a noise term with mean m and variance σ^2 . The agent receives a wage $\alpha + \beta \cdot \pi$. For simplicity assume that the agent is risk neutral and maximizes her utility

$$\alpha + \beta \cdot \pi - c(e).$$

Suppose now that the principal can also introduce a monitoring activity (performance review) $r \in \{0,1\}$ carried out by the agent's respective supervisor. The agent anticipates that she will incur psychological or economic costs from "underperforming" some effort level $\hat{e} > \bar{e}$ when performance reviews are in place (where \hat{e} may, for instance, be the first-best effort level). If reviews are in place (i.e. $r = 1$) her utility is reduced by $g(\hat{e} - e)$ where $g(\Delta) = 0$ for $\Delta \leq 0$ and $g' > 0, g'' \geq 0$ for $\Delta > 0$. The agent then maximizes

$$\max_e \beta(e + m) - c(e) - r \cdot g(\hat{e} - e).$$

Consider now a situation where initially there is neither performance pay nor are performance reviews used such that initially the agent chooses $e = \bar{e}$. Relative to this, both instruments naturally raise performance but are substitutes as the following results show:

Proposition 1: *The introduction of performance pay and performance reviews both raise performance. Both instruments are substitutes: The introduction of performance reviews has a weaker additional effect on performance if performance pay is in place.*

Proof: The first derivative of the agent's objective function is

$$\begin{cases} \beta - c'(e) + rg'(\hat{e} - e) & \text{if } e < \bar{e} \\ \beta - c'(e) & \text{if } e \geq \bar{e}. \end{cases}$$

Without performance reviews and performance pay the agent chooses $e = \bar{e}$. If $c'^{-1}(\beta) \geq \hat{e}$ (i.e. if β is sufficiently large) then the agent always chooses $e = c'^{-1}(\beta)$ irrespective of the monitoring activity. If this is not the case, optimal efforts are characterized by

$$\beta - c'(e) + rg'(\hat{e} - e) = 0.$$

Hence, efforts are in this case increasing in r : when performance reviews are in place, marginal returns to efforts are higher, as higher efforts (below \hat{e}) then additionally reduce the psychological costs of underperformance. In this case by the implicit function theorem we have that

$$\frac{\partial e}{\partial \beta} = \frac{1}{c''(e) + rg''(\hat{e} - e)}$$

which implies that

$$\left. \frac{\partial e}{\partial \beta} \right|_{r=0} > \left. \frac{\partial e}{\partial \beta} \right|_{r=1} > 0.$$

■

7.2. A Model of Performance Pay and Self-Reliance

Consider a simple formal model. An agent A exerts effort a_i on two tasks $i \in \{1,2\}$ at personal costs $\frac{c_i}{2}a_i^2$. The agent's work generates a profit $a_1 + a_2$ for the firm and earns a wage $w = \alpha + \beta(a_1 + a_2)$. The agent interacts with a supervisor S . While task 1 only affects the agent, task 2 generates an externality on the supervisor (for instance has it requires the supervisor's help or imposes additional workload on her), reducing the supervisor's utility by xa_2 . Hence, first best actions are determined by

$$\max_{a_1, a_2} a_1 + (1-x)a_2 - \frac{c_1}{2}a_1^2 - \frac{c_2}{2}a_2^2$$

and thus given by

$$\begin{aligned} a_1^{FB} &= \frac{1}{c_1} \text{ and} \\ a_2^{FB} &= \frac{1-x}{c_2}. \end{aligned}$$

The agent's utility depends on her income and costs of effort but he also cares for overall surplus $a_1 + (1-x)a_2$ – thus internalizing the effect of his actions on her supervisor to some extent. The agent exhibits image concerns as in Bénabou/Tirole (2006) and we assume that he wants to be viewed as someone who cares for the overall surplus. The agent's utility function is thus

$$\gamma(\alpha + \beta(a_1 + a_2)) + \eta(a_1 + (1-x)a_2) - \frac{c_1}{2}a_1^2 - \frac{c_2}{2}a_2^2 + v_\eta E[\eta|a_1, a_2],$$

where γ measures the weight the agent puts on monetary income and η the weight on overall surplus. Both weighting parameters are unknown to the supervisor and drawn from a continuous

joint distribution on \mathbb{R}^2 and $E[\eta|a_1, a_2]$ denotes the supervisor's posterior expectation about η . Moreover, v_η represents the degree of the agent's image concerns regarding η .

We here focus on equilibria in which these expectations are linear and additively separable in a_1 and a_2 such that $E[\eta|a_1, a_2] = \lambda_0 + \lambda_1 a_1 + \lambda_2 a_2$. We show the existence of such linear equilibria and characterize equilibrium efforts and profits. We start with the case in which there is no bonus, i.e. $\beta = 0$.⁴⁸

Proposition 1. *When $\beta = 0$ there is a continuum of equilibria which all generate the same gross surplus $a_1 + (1 - x)a_2 = \frac{\eta}{c_1} + \frac{\eta(1-x)^2}{c_2} + v_\eta$. In the net surplus maximizing equilibrium efforts are*

$$a_1 = \frac{\eta}{c_1} + v_\eta \frac{1}{1 + \frac{c_1}{c_2}(1-x)^2} \text{ and}$$

$$a_2 = \frac{\eta(1-x)}{c_2} + v_\eta \frac{\frac{c_1}{c_2}(1-x)}{1 + \frac{c_1}{c_2}(1-x)^2}.$$

and stronger image concerns v_η raise efforts for both tasks and thus overall profits.

Image concerns here can thus raise efforts and profits: The agent works harder to signal that he cares for the overall surplus. And higher efforts for both tasks can serve to signal this motivation. When performance reviews trigger image concerns they therefore increase profits and surplus.

However, this result changes when a bonus is implemented and $\beta > 0$ as the following result shows:

Proposition 2. *When $\beta > 0$ the agent chooses*

$$a_1 = \frac{\gamma\beta + \eta}{c_1} + \frac{v_\eta}{x} \text{ and}$$

$$a_2 = \frac{\gamma\beta + \eta(1-x)}{c_2} - \frac{v_\eta}{x}.$$

Stronger image concerns v_η raise efforts for task 1 but reduce efforts for task 2 and have no effect on profits.

⁴⁸ Proofs are provided at the end of this section.

Once a bonus is implemented, higher image concerns raise the efforts for task 1 on which the agent works individually but reduce efforts for task 2 which involves the supervisor. The reason is the following: Both tasks 1 and 2 have the same impact on profits (and thus the agent's bonus payment), but task 2 imposes an externality on the supervisor. When there is no bonus both efforts serve as signals of the agent's motivation to raise surplus. If, however, the agent receives a bonus, higher efforts for task 2 relative to those exerted on task 1 indicate that the agent cares more for his own bonus rather than overall surplus. By exerting relatively less effort on task 2 and more on task 1 the agent thus can signal that she is more interested in raising the surplus (and thus also internalizing the effect of her actions on the supervisor) than in mere bonus payments. This drives a wedge between the signaling value of efforts for both tasks.

In other words, the agent wants to avoid giving the impression that he wants to enrich himself at the supervisor's expense and will do so by lowering efforts for task 2. Performance gains for task 1 are offset by performance losses for task 2. Hence, when performance reviews trigger image concerns they here have no effect on profits.

Moreover, the introduction of a bonus leads to a reduction of efforts for the task which imposes an externality on the supervisor.

Proof of Proposition 1:

Given the observers' beliefs the agent maximizes

$$\max_{a_1, a_2} \eta(a_1 + (1-x)a_2) - \frac{c_1}{2} a_1^2 - \frac{c_2}{2} a_2^2 + v_\eta(\lambda_0 + \lambda_1 a_1 + \lambda_2 a_2).$$

From the first order conditions we obtain the best response

$$\begin{aligned} a_1 &= \frac{\eta + v_\eta \lambda_1}{c_1} \text{ and} \\ a_2 &= \frac{\eta(1-x) + v_\eta \lambda_2}{c_2}. \end{aligned} \tag{1}$$

For this to be an equilibrium we must have that beliefs are correct and thus

$$\begin{aligned} \eta &= \lambda_0 + \lambda_1 \frac{\eta + v_\eta \lambda_1}{c_1} + \lambda_2 \frac{\eta(1-x) + v_\eta \lambda_2}{c_2} \\ &= \lambda_0 + \left(\frac{\lambda_1}{c_1} + \lambda_2 \frac{(1-x)}{c_2} \right) \eta + \frac{v_\eta \lambda_1^2}{c_1} + \frac{v_\eta \lambda_2^2}{c_2} \end{aligned}$$

must hold for all $\eta \in \mathbb{R}$ which requires that

$$\lambda_0 = -\frac{v_\eta \lambda_1^2}{c_1} - \frac{v_\eta \lambda_2^2}{c_2} \quad \text{and} \quad \frac{\lambda_1}{c_1} + \lambda_2 \frac{(1-x)}{c_2} = 1.$$

From the latter we must have that

$$\lambda_2 = \frac{c_2}{1-x} \left(1 - \frac{\lambda_1}{c_1}\right).$$

Hence, there is a continuum of linear equilibria indexed by λ_1 where

$$E[\eta|a_1, a_2] = \lambda_0 + \lambda_1 a_1 + \frac{c_2}{1-x} \left(1 - \frac{\lambda_1}{c_1}\right) a_2.$$

Substituting efforts best responses as given by (1) we obtain the gross surplus

$$a_1 + (1-x)a_2 = \frac{\eta}{c_1} + \frac{\eta(1-x)^2}{c_2} + v_\eta$$

which is thus identical in all equilibria. The net surplus maximizing equilibrium is determined by

$$\min_{\lambda_1} \frac{c_1}{2} \left(\frac{\eta + v_\eta \lambda_1}{c_1}\right)^2 + \frac{c_2}{2} \left(\frac{\eta(1-x)}{c_2} + \frac{v_\eta}{1-x} \left(1 - \frac{\lambda_1}{c_1}\right)\right)^2$$

yielding

$$\lambda_1 = \frac{1}{\frac{1}{c_1} + \frac{(1-x)^2}{c_2}}.$$

Substituting this for λ_1 and $\lambda_2 = \frac{c_2}{1-x} \left(1 - \frac{\lambda_1}{c_1}\right)$ respectively into the agent's reaction functions yields the equilibrium efforts. ■

Proof of Proposition 2:

The agent maximizes

$$\gamma\beta(a_1 + a_2) + \eta(a_1 + (1-x)a_2) - \frac{c_1}{2} a_1^2 - \frac{c_2}{2} a_2^2 + v_\eta(\lambda_0 + \lambda_1 a_1 + \lambda_2 a_2)$$

and from the first order conditions we obtain

$$\begin{aligned} a_1 &= \frac{\gamma\beta + \eta + v_\eta \lambda_1}{c_1} \quad \text{and} \\ a_2 &= \frac{\gamma\beta + \eta(1-x) + v_\eta \lambda_2}{c_2}. \end{aligned}$$

Moreover,

$$\begin{aligned}\eta &= c_1 a_1 - v_\eta \lambda_1 - \gamma \beta \text{ and} \\ \gamma &= \frac{c_2 a_2 - v_\eta \lambda_2 - \eta(1-x)}{\beta}.\end{aligned}$$

From solving this system of two equations we obtain that

$$\eta = \frac{c_1 a_1 - c_2 a_2 + v_\eta (\lambda_2 - \lambda_1)}{x}.$$

Therefore $\lambda_1 = \frac{c_1}{x}$ and $\lambda_2 = -\frac{c_2}{x}$ such that beliefs are then indeed linear. Hence, such an equilibrium exists and equilibrium efforts are obtained from substituting λ_1 and λ_2 into the agent's reaction functions. Note furthermore that profits

$$a_1 + a_2 = \frac{\gamma \beta + \eta}{c_1} + \frac{\gamma \beta + \eta(1-x)}{c_2}.$$

are independent of v_η . ■

7.3. Tables and Figures

Table A1: Classification of Store Manager Tasks

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

Table A2: Balancing Table

	(1)	(2)	(3)	(4)	(5)
	Descriptives Overall	Descriptives Control	Descriptives Bonus	Descriptives Review	Descriptives Bonus&Review
Profits Jan-Mar '17	26511.69 (10963.98)	27776.48 (11949.38)	25549.85 (11373.97)	26138.93 (8450.81)	26381.56 (11544.57)
Planned Profits Jan-Mar '17	28166.79 (10229.54)	28827.52 (11253.13)	28221.86 (10796.1)	27397.36 (8226.94)	28103.59 (10367.23)
Planned Profits Apr-Jun '17	28288.69 (10299.1)	28979.56 (11294.37)	28392.91 (10795.42)	27334.64 (7999.76)	28303.54 (10693.81)
Female Store Manager (Y/N)	0.55 (0.50)	0.67 (0.48)	0.68 (0.47)	0.4*** (0.50)	0.44** (0.50)
Walking Customers (Y/N)	0.13 (0.34)	0.12 (0.03)	0.14 (0.35)	0.16 (0.37)	0.13 (0.34)
FTE	6.39 (1.35)	6.54 (1.10)	6.23 (1.49)	6.22 (1.27)	6.52 (1.51)
Age of Store	14.90 (8.79)	13.35 (8.18)	14.09 (9.39)	17.65** (9.19)	14.85 (10.05)
Age Store Manager	41.42 (9.61)	41.93 (9.78)	42.49 (9.75)	39.5 (9.02)	41.60 (10.05)
Tenure Store Manager	15.33 (8.86)	15.80 (8.84)	16.90 (8.09)	13.55 (8.61)	14.98 (9.59)
Tenure District Manager	13.12 (11.05)	14.82 (10.31)	10.10 (10.11)	15.04 (11.00)	12.47 (12.23)
Store Space	710.22 (145.53)	744.57 (134.88)	714.29 (143.39)	689.08 (179.76)	691 (121.86)
Max. Observations	224	60	51	50	63

Note: The table reports means of the respective variables for the different treatment groups and their standard deviations in parentheses. Asterisks display significance levels from t-tests (fisher exact test for binary variables) of the respective treatment group against the control group. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Regression including only Treatment Months

	(1)	(2)	(3)	(4)
	Profits	log (Profits)	Profits	log (Profits)
Treatment Effect BONUS	-366.59 (580.88)	-0.0502 (0.0365)	-302.60 (627.39)	-0.0357 (0.0404)
Treatment Effect REVIEW	1101.66** (514.86)	0.0650** (0.0296)	1390.47** (534.66)	0.0649** (0.0262)
Treatment Effect BONUS&REVIEW	-733.38 (492.00)	-0.0202 (0.0297)	-638.23 (523.46)	-0.0215 (0.0286)
Wald test REVIEW=BONUS&REVIEW	$p=0.0002$	$p=0.0065$	$p=0.0001$	$p=0.0041$
Time FE	Yes	Yes	Yes	Yes
Refurbishments	Yes	Yes	Yes	Yes
Planned Profits	Yes	Yes	Yes	Yes
Further Controls	No	No	Yes	Yes
N Observations	669	669	669	669
N Stores	224	224	224	224
N Cluster	31	31	31	31
Overall R^2	0.8696	0.6491	0.8726	0.6622

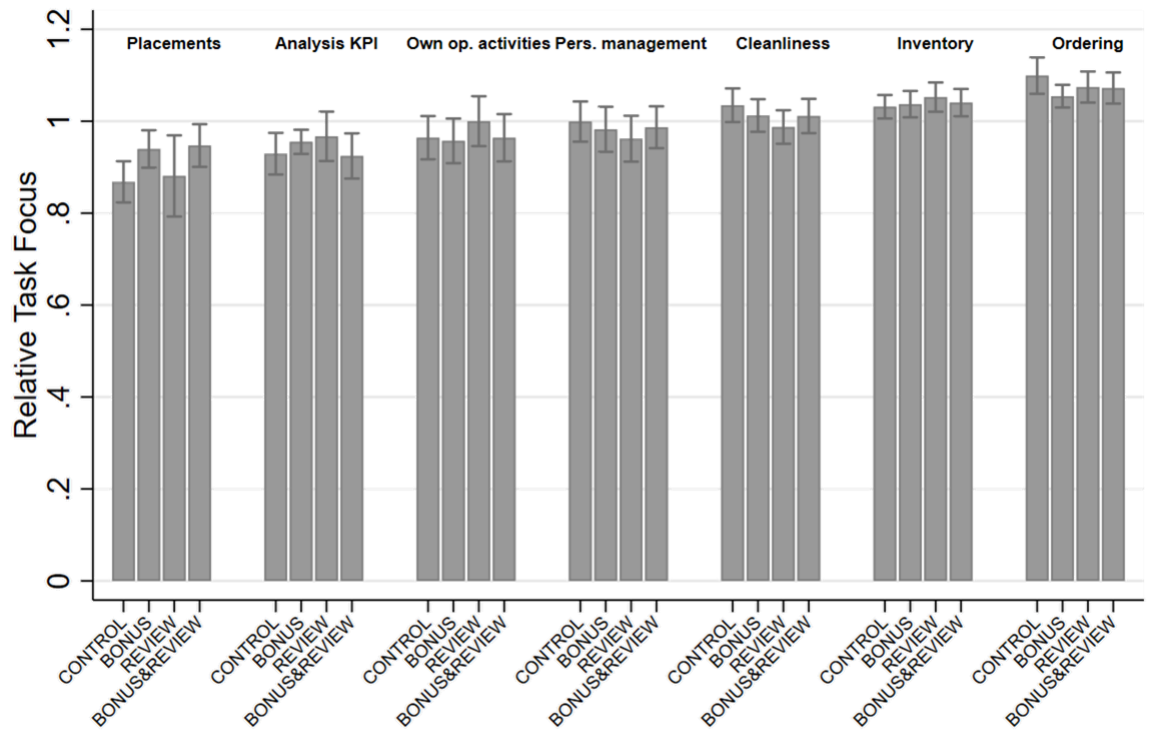
Note: The table reports results from ordinary least squares regressions using only data from the treatment period further controlling for the mean of profits from January 2016-March 2017. All regressions control for possible refurbishments of a store, the randomization pair, and the companies planned profits. Columns 3&4 further control for variables with slight imbalance between treatments (gender, age of the store). Observations are excluded once a store manager switched the store during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A4: Monthly Treatment Effects

	(1)	(2)	(3)	(4)
	Profits	Profits	Log (Profits)	Log (Profits)
Treatment Effect BONUS	-74.22	293.0	-0.0181	0.00471
1 st Month	(637.2)	(684.7)	(0.0325)	(0.0429)
Treatment Effect BONUS	572.8	912.4	0.0327	0.0536*
2 nd Month	(726.4)	(820.0)	(0.0314)	(0.0314)
Treatment Effect BONUS	-585.1	-202.9	-0.0237	-0.000987
3 rd Month	(928.8)	(1053.1)	(0.0941)	(0.110)
Treatment Effect BONUS	-1379.8	-1014.3	-0.0554	-0.0381
4 th Month (after treatment)	(1032.5)	(1074.5)	(0.0449)	(0.0514)
Treatment Effect BONUS	854.5	1225.3	-0.0196	-0.00470
5 th Month (after treatment)	(1436.0)	(1626.6)	(0.0271)	(0.0402)
Treatment Effect REVIEW	1417.1*	1465.3	0.0645*	0.0751*
1 st Month	(783.4)	(867.4)	(0.0332)	(0.0442)
Treatment Effect REVIEW	2451.7***	2490.4***	0.0957***	0.104***
2 nd Month	(618.8)	(692.7)	(0.0291)	(0.0295)
Treatment Effect REVIEW	461.9	966.1	0.0680	0.0922
3 rd Month	(782.6)	(889.9)	(0.0551)	(0.0723)
Treatment Effect REVIEW	-1038.2	-461.8	-0.0599	-0.0332
4 th Month (after treatment)	(1149.4)	(1255.2)	(0.0493)	(0.0592)
Treatment Effect REVIEW	746.8	1086.4	0.0205	0.0342
5 th Month (after treatment)	(685.7)	(1044.9)	(0.0255)	(0.0433)
Treatment Effect BONUS&REVIEW	-590.7	-474.2	-0.0274	-0.0184
1 st Month	(590.5)	(511.4)	(0.0294)	(0.0398)
Treatment Effect BONUS&REVIEW	801.1	886.8	0.0267	0.0306
2 nd Month	(686.1)	(664.7)	(0.0364)	(0.0381)
Treatment Effect BONUS&REVIEW	-1074.2	-958.3	-0.00156	0.000577
3 rd Month	(1030.9)	(1165.6)	(0.0751)	(0.0970)
Treatment Effect BONUS&REVIEW	-656.0	-456.7	-0.0536	-0.0513
4 th Month (after treatment)	(1234.0)	(1410.9)	(0.0589)	(0.0768)
Treatment Effect BONUS&REVIEW	-121.7	-30.55	-0.0260	-0.0297
5 th Month (after treatment)	(709.6)	(940.2)	(0.0349)	(0.0507)
Fixed Effects (Time, Store)	Yes	Yes	Yes	Yes
Fixed Effects (District & Store Manager)	No	Yes	No	Yes
Refurbishments	Yes	Yes	Yes	Yes
Planned Profits	Yes	Yes	Yes	Yes
Observations	4421	4203	4412	4194
N Store	224	224	224	224
N Cluster	31	31	31	31
Within R^2	0.2407	0.2726	0.1703	0.1938
Overall R^2	0.7484	0.5379	0.6152	0.4063

Note: The table reports results from fixed effects regressions with the profits on the store level as the dependent variable. The regression accounts for time and store fixed effects (column 1-4) and adds fixed effects for district and store managers in columns 2&4. The fixed effects regressions compare pre-treatment observations (January 2016-March 2017) with the observations during the experiment (April 2017 – June 2017). All regressions control for possible refurbishments of a store and the companies planned value. Observations are excluded once a store manager switched the store during the treatment period. *Treatment Effect* thus refers to the difference-in-difference estimator. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A1: Relative Task Focus (Post-experimental survey)



Note: The figure displays the average rating of focus on specific tasks (1=low focus, 6=high focus) obtained from an online questionnaire. Tasks were clustered into 7 dimensions. The average focus of a dimension was then divided by the average focus of all dimensions. 95% confidence bars are displayed.

Table A5: Main Treatment Effects on Profit Components

	(1)	(2)	(3)	(4)	(5)	(6)
	Gross Profit Margin	Personnel Costs	Inventory Losses	Gross Profit Margin	Personnel Costs	Inventory Losses
Treatment Effect BONUS	-0.261 (0.337)	-0.397 (0.516)	0.0236 (0.286)	0.369 (0.353)	-0.246 (0.560)	0.313* (0.176)
Treatment Effect REVIEW	0.282 (0.523)	-0.677* (0.394)	-0.336 (0.351)	1.111** (0.541)	-0.362 (0.324)	-0.141 (0.252)
Treatment Effect BONUS&REVIEW	-0.976* (0.546)	-0.703** (0.318)	-0.0490 (0.252)	-0.550 (0.431)	-0.741*** (0.245)	0.330 (0.214)
Wald test REVIEW=BONUS&REVIEW	$p=0.0531$	$p=0.9449$	$p=0.3502$	$p=0.0019$	$p=0.1995$	$p=0.1028$
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes	Yes
District Manager FE	No	No	No	Yes	Yes	Yes
Store Manager FE	No	No	No	Yes	Yes	Yes
Refurbishments	Yes	Yes	Yes	Yes	Yes	Yes
Planned Profits	Yes	Yes	Yes	Yes	Yes	Yes
N of Observations	3975	3977	3895	3776	3779	3701
N of Stores	224	224	224	224	224	224
Cluster	31	31	31	31	31	31
Within R^2	0.5525	0.2438	0.1377	0.5631	0.2843	0.1903
Overall R^2	0.9197	0.6538	0.0694	0.8918	0.3795	0.1903

Note: The table reports results from a fixed effects regression with the different profit components on the store level as the dependent variable. The regression accounts for time and store fixed effects and adds fixed effects for district manager and store managers in column 4-6. The regressions compare pre-treatment observations (January 2016 - March 2017) with the observations during the experiment (April 2017 – June 2017). *Treatment Effect* thus refers to the difference-in-difference estimator. All regressions control for possible refurbishments of a store and the companies planned value of the respective profit components. Observations are excluded when a store manager switched the store during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Content of Review Meetings

	(1)	(2)
<i>Reference Group:</i>		
<i>Treatment REVIEW</i>		
Ordering	0.0898*** (0.0182)	0.0924*** (0.0146)
Placements	0.0983*** (0.0131)	0.0994*** (0.0141)
Cleanliness	0.0386* (0.0220)	0.0332 (0.0250)
Analysis KPI	0.00368 (0.0167)	-0.00450 (0.0174)
Inventory	0.0507*** (0.0169)	0.0514*** (0.0170)
BONUS x Ordering	-0.0353* (0.0198)	-0.0308* (0.0173)
BONUS x Placements	0.0210 (0.0136)	0.0305* (0.0159)
BONUS x Cleanliness	-0.0432** (0.0191)	-0.0367** (0.0167)
BONUS x Analysis KPI	0.0140 (0.0229)	0.0267 (0.0257)
BONUS x Inventory	-0.0127 (0.0179)	-0.0121 (0.0170)
Planned task (section 2)	-0.00420 (0.0128)	-0.00500 (0.0129)
Problems encountered (section 3)	-0.0685*** (0.0102)	-0.0670*** (0.00860)
Meeting slot 2	0.00438 (0.00562)	0.00376 (0.00539)
Meeting slot 3	0.00746 (0.00815)	0.000295 (0.00809)
Meeting slot 4	-0.000128 (0.00590)	0.00165 (0.00717)
Meeting slot 5	-0.0107 (0.0135)	-0.00810 (0.0152)
Meeting slot 6	-0.000548 (0.00976)	-0.00765 (0.00909)
Controls	No	Yes
Observations	35931	27318
Cluster	18	17
Pseudo R^2	0.1377	0.1492

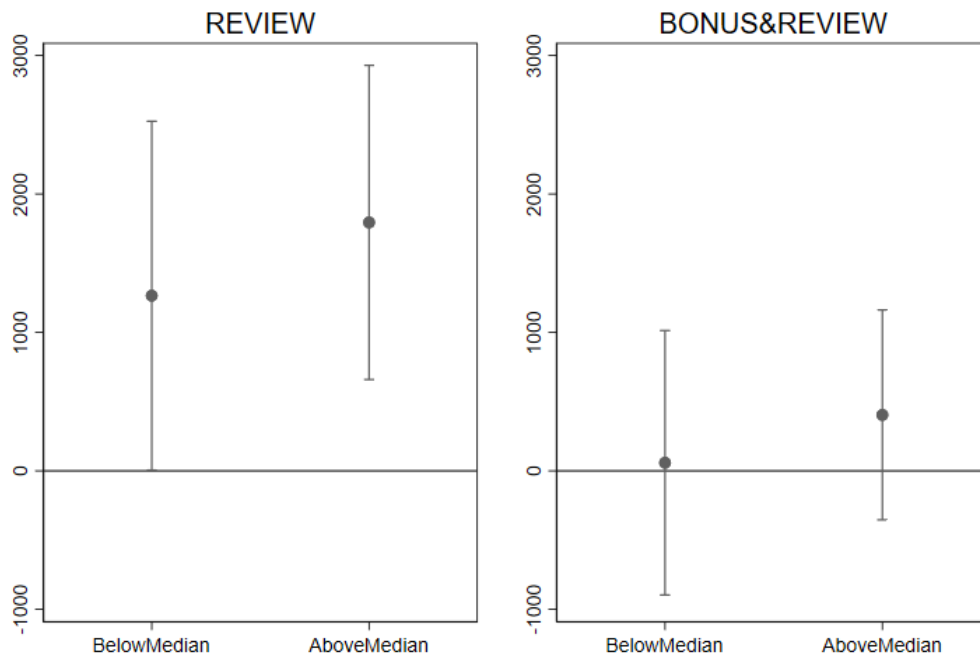
Note: The table reports results from Probit regressions. Dependent variable y_{kst} is a dummy variable indicating whether a task k was mentioned in section s of a review meeting conducted in time slot t . Further controls in columns (4)-(6) are store size, number of employees, store manager's age and prior performance evaluation, as well as randomization group. The Treatment REVIEW serves as the reference group. Standard errors are clustered on the district level at treatment start and displayed in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Performance Evaluations and Future Wages

	(1) Log Wage	(2) Log Wage
Perf. Eval.=1 (low performer)	<i>Reference Group</i>	
Perf. Eval.=2	0.0259** (0.0121)	0.0280** (0.0115)
Perf. Eval.=3	0.0511*** (0.0113)	0.0515*** (0.0108)
Perf. Eval.=4 (high performer)	0.0808*** (0.0151)	0.0778*** (0.0149)
Tenure		0.0014*** (0.0003)
Store Space		0.0001 (0.0001)
N of Observations	764	764
<i>Overall R²</i>	0.0468	0.0755

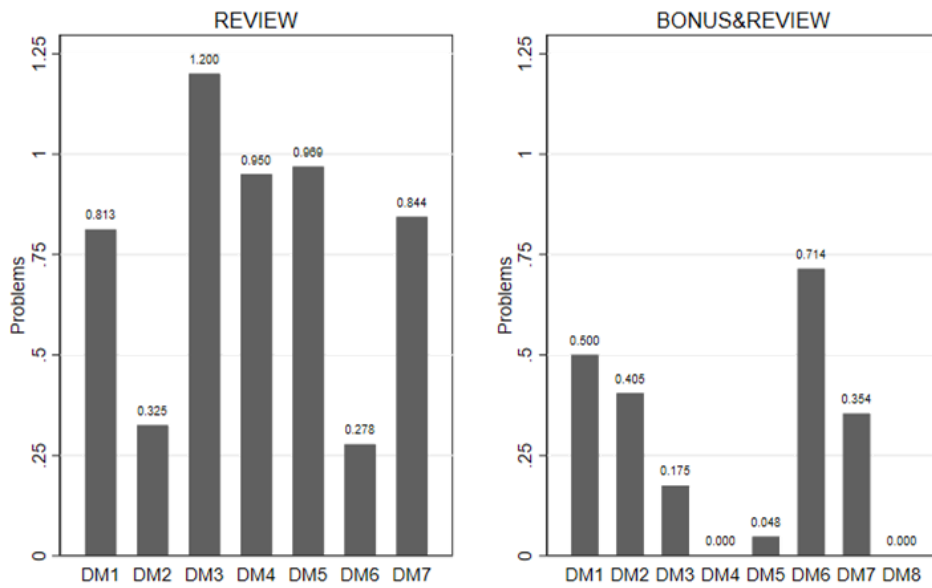
Note: The table reports results from OLS regressions with log monthly wages of store managers in 2018 as the dependent variable. *Perf. Eval.* is a set of dummy variables and refers to the store managers' annually made subjective performance evaluation of supervisors (district managers) in 2016 with 1=low performer and 4=high performer. *Perf. Eval.*=0 (the lowest group) is the reference group and thus omitted. Robust standard errors are displayed in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A2: Average Treatment Effects Depending on the Number of Review Conversations



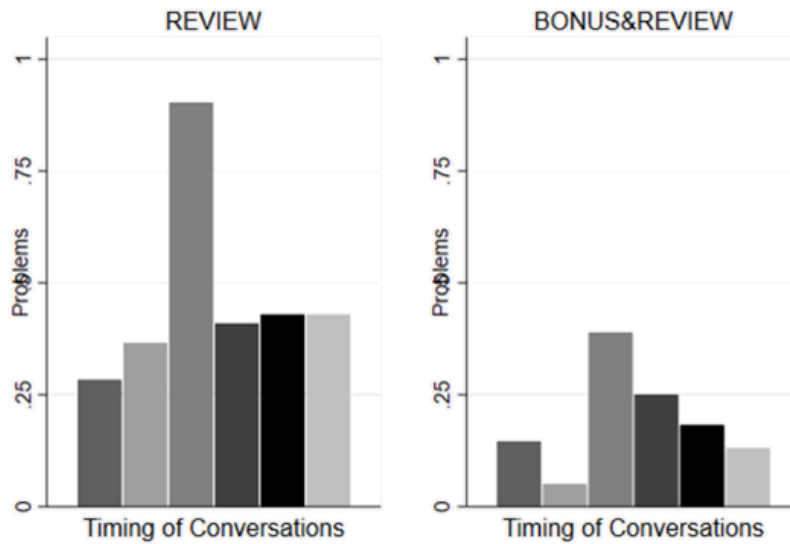
Note: The figure displays separately estimated treatment effects from our standard fixed effects regression specification depending on whether the number of performance reviews conducted is below or above/equal to the median (4). 95% confidence bars are displayed.

Figure A3: Average Number of Notes in Subsection “Problems” per Conversation by District Manager



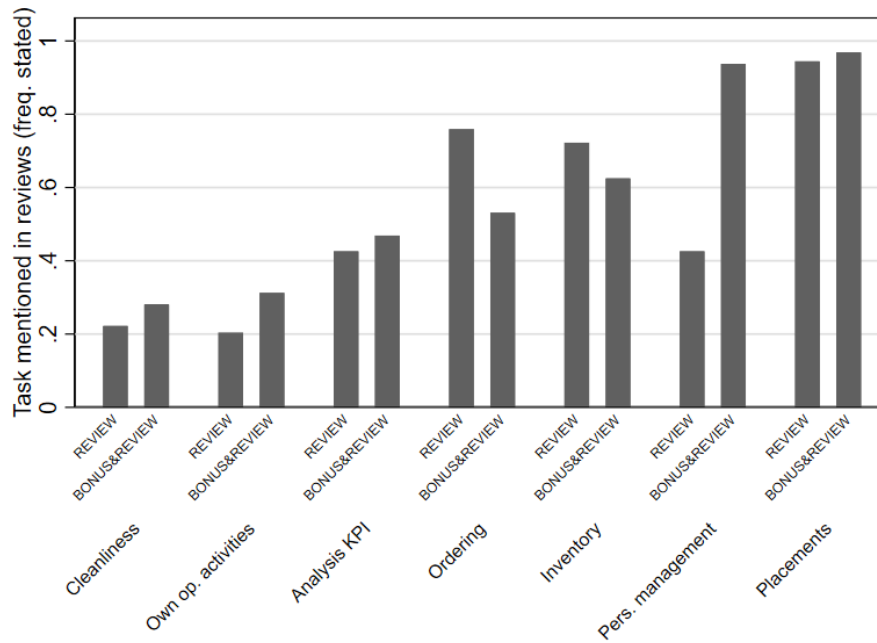
Note: The figure displays the average number of problems (in notes/sentences) per session displayed for each district manager (DM) separately.

Figure A4: Average Number of Notes in Subsection “Problems” per Conversation Depending on Time of Conversation



Note: The figure displays the average number of problems (in notes/sentences) per session. The average number of problems is displayed for each time point separately (1st bar= 1st two weeks, 2nd bar= 2nd two weeks, etc.).

Figure A5: Tasks covered in review conversations



Note: The figure displays the frequency of mentioned categories per review conversation (all categories – tasks done, problems, tasks next time - pooled).

7.4. Instructions (Online Appendix)

7.4.1. Store Manager - CONTROL Group (sent to their home address, originally in German)

Project DB1⁴⁹

Dear Mr./Mrs. XXX,

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

You will now have access to the information package.

Learning Unit⁵⁰:

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

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

Access data learning unit:

Please visit the following website for the learning unit:

Your password is:

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

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

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

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

Yours sincerely

⁴⁹ The company uses „DB1“ (short for Deckungsbeitrag 1/ contribution margin) as an internal title for the simplified profit measure explained above in our study: $Profit = Net\ Sales - Cost\ of\ Goods\ Sold - Staff\ Costs - Inventory\ Losses$

⁵⁰ Due to previous company wording, the company uses “learning unit” as an internal description for the learning video, the quiz, the margin information and the monthly feedback. We refer to this as “information package” in the above.

7.4.2. Store Manager – BONUS Group (send to their home address, originally in German)

Project DB1

Dear Mr./Mrs. XXX,

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

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

Bonus:

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

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

Calculation: DB1-Bonus (in €) = (DB1 – 80% of the Plan DB1) * 0,05

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

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

Learning Unit:

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

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

Access data learning unit:

Please visit the following website for the learning unit:

Your password is:

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

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

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

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

Yours sincerely

7.4.3. Store Manager – REVIEW Group (send to their home address, originally in German)

Project DB1

Dear Mr./Mrs. XXX,

a positive DB1 profit measure is important for the economic success of [*the company*]. For this reason, the DB1 project will be implemented in your region during the next few months. Within the scope of the DB1 project, you will receive a learning unit and have a regular DB1-Conversation with your district manager in the near future.

You will now have access to the learning unit. Your district manager will contact you regarding the DB1-Conversation.

Learning Unit:

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

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

Access data learning unit:

Please visit the following website for the learning unit:

Your password is:

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

District manager DB1-Conversation:

Your district manager will also have an in-depth DB1-Conversation with you every two weeks. Within this conversation, he will ask you about actions you have already taken to increase the DB1 profit measure. In addition, you can discuss possible problems with him.

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

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

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

Yours sincerely

7.4.4. Store Manager – BONUS&REVIEW Group (send to their home address, originally in German)

Project DB1

Dear Mr./Mrs. XXX,

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

Your bonus period starts on 01.04.2017 for 3 months. You will now have access to the learning unit. Your district manager will contact you regarding the DB1-Conversation.

Bonus:

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

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

Calculation: DB1-Bonus (in €) = (DB1 – 80% of the Plan DB1) * 0,05

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

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

Learning Unit:

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

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

Access data learning unit:

Please visit the following website for the learning unit:

Your password is:

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

District manager DB1-Conversation:

Your district manager will also have an in-depth DB1-Conversation with you every two weeks. Within this conversation, he will ask you about actions you have already taken to increase the DB1 profit measure. In addition, you can discuss possible problems with him.

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

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

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

Yours sincerely

7.4.5. Information about the DB1-Bonus (added to both BONUS treatments)

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

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

Calculation DB1-Bonus

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

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

Amount in euros = $(\text{DB1} - 80\% \text{ plan DB1}) * 0,05$

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

Fictitious Example

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

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

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

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

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

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

Total bonus paid: $380 \text{ (April)} - 60 \text{ (May)} + 240 \text{ (June)} = 560\text{€}$

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

7.4.6. Monthly Communication to Store Manager (sent to their home address, originally in German)

Project DB1

Dear Mr./Mrs. XXX,

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

Summary of your DB1 profit measure in April 2017:⁵¹

(Amounts are not rounded until the end)

Sales:

Cost of good sold:

Personnel costs:

Inventory:

This results in a DB1 April/2017:

For a plan DB1 April/2017:

The resulting bonus amount for the month of April is:

$(DB1 - 0.8 * \text{plan DB1}) * 0.05 \text{ €} =$

Summary of your bonus amounts since April 2017:

(Amounts are not rounded until the end)

Bonus amount April 2017: € (gross)

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

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

⁵¹ For accounting reasons, the letter in May came with additional information: "In April, adjusting entries through accounting were posted to the region only and not distributed to the branches. Their profit margin is therefore too well represented. These bookings will be made up with the May-finalization. Therefore, you will find the margin correction in your May letter with a reversed sign. In sum of April and May, the correction value will be € 0.00. We ask for your understanding."

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

Project DB1

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

Within the scope of the DB1 project, all store managers will participate in a learning unit about the DB1 profit measure in the near future. In addition, stores in randomly selected districts receive an additional DB1-Conversation. Moreover, an additional bonus for store managers is introduced in all stores of the region. For administrative reasons, the bonus will be introduced in the districts at different times. The assignment happens randomly according to a statistical procedure

In your district, store managers will receive the bonus and the additional Profit-Conversation at a later date. You will be informed in sufficient time about the exact time frame. However, store managers will have access to an online learning unit with regard to the DB1 profit measure from 27.03.2017. Please make sure that the learning unit is completed by the store managers in your district.

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

Store manager learning unit:

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

Your personal password is: XXXXX

Communication upon inquiries of store managers:

- If your store managers ask why you do not have a Profit-Conversation with them, we ask that you communicate that this is a random selection and that your area's store managers will have such a conversation at a later time.
- If your store managers ask why they are not getting a bonus for the increased DB1 profit measure, we also ask you to communicate that this is a random selection and that the store managers in your district will in any case receive a bonus at a later date.

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

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

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

Yours sincerely

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

Project DB1

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

Within the scope of the DB1 project, all store managers will participate in a learning unit about the DB1 profit measure in the near future. In addition, stores in randomly selected districts receive an additional DB1-Conversation. Moreover, an additional bonus for store managers is introduced in all stores of the region. For administrative reasons, the bonus will be introduced in the districts at different times. The assignment happens randomly according to a statistical procedure

From 01.04.2017, store managers in your district will be given the opportunity to receive a bonus for 3 months, but they will not receive an additional Profit-Conversation for the time being. Store managers will also have access to an online learning unit regarding DB1 profit measure from 27.03.2017. Please make sure that the learning unit is completed by the store managers in your district.

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

Store manager bonus:

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

Calculation: DB1-Bonus (in €) = (DB1 – 80% of the Plan DB1) * 0,05

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

Store manager learning unit:

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

Your personal password is: XXXXX

Communication upon inquiries of store managers:

If your store managers ask why you do not have a Profit-Conversation with them, we ask that you communicate that this is a random selection and that your area's store managers will have such a conversation at a later time.

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

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

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

Yours sincerely

7.4.9. District Manager – REVIEW Group (sent to their e-mail address, originally in German)

Project DB1

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

Within the scope of the DB1 project, all store managers will participate in a learning unit about the DB1 profit measure in the near future. In addition, stores in randomly selected districts receive an additional DB1-Conversation. Moreover, an additional bonus for store managers is introduced in all stores of the region. For administrative reasons, the bonus will be introduced in the districts at different times. The assignment happens randomly according to a statistical procedure

From 27.03.2017, store managers will have access to a learning unit regarding the DB1 profit measure. Please make sure that the learning unit is completed by the store managers in your district.

From 01.04.2017, an additional DB1-Conversation will be introduced in your district. In your district, store managers will receive the DB1-Bonus at a later date. You will be informed in sufficient time about the exact time frame.

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

Your store manager DB1-Conversation:

We would like to ask you to hold an in-depth personal conversation with the store managers in your district every two weeks about the development of the DB1 profit measure (DB1-Conversation).

For this DB1-Conversation, we have attached a guideline for you which we would like you to fill out in note form with every conversation and send it back to your personnel management. During your conversation, you should not only inquire and examine what the store manager did, but also communicate what they should do differently until the next meeting. The DB1-Conversation should happen every two weeks on the key dates 18.04.2017, 02.05.2017, 16.05.2017, 30.05.2017, 13.06.2017, 27.06.2017. Store managers will be informed individually in a separate letter.

Store manager learning unit:

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

Your personal password is: XXXXX

Communication upon inquiries of store managers:

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

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

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

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

Yours sincerely

Conversation guideline

Key date: 18.04.2017 2.05.2017 16.5.2017 30.05.2017 13.06.2017 27.06.2017

Store Manager:

What has the store manager done to increase the DB1?

What problems have occurred?

Which measures / which next steps does the store manager want to carry out until the next meeting?

7.4.10. District Manager – BONUS&REVIEW Group (sent to their e-mail address, originally in German)

Project DB1

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

Within the scope of the DB1 project, all store managers will participate in a learning unit about the DB1 profit measure in the near future. In addition, stores in randomly selected districts receive an additional DB1-Conversation. Moreover, an additional bonus for store managers is introduced in all stores of the region. For administrative reasons, the bonus will be introduced in the districts at different times. The assignment happens randomly according to a statistical procedure

From 01.04.2017, store managers in your district will be given the opportunity to receive a bonus for 3 months and an additional Profit-Conversation will be introduced. Store managers will also have access to an online learning unit regarding the contribution margin 1 from 27.03.2017. Please make sure that the learning unit is completed by the store managers in your district.

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

Store manager bonus:

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

Calculation: DB1-Bonus (in €) = (DB1 – 80% of the Plan DB1) * 0,05

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

Your store manager DB1-Conversation:

We would like to ask you to hold an in-depth personal conversation with the store managers in your district every two weeks about the development of the DB1 profit measure (DB1-Conversation).

For this DB1-Conversation, we have attached a guideline for you which we would like you to fill out in note form with every conversation and send it back to your personnel management. During your conversation, you should not only inquire and examine what the store manager did, but also communicate what they should do differently until the next meeting. The DB1-Conversation should happen every two weeks on the key dates 18.04.2017, 02.05.2017, 16.05.2017, 30.05.2017, 13.06.2017, 27.06.2017. Store managers will be informed individually in a separate letter.

Store manager learning unit:

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

Your personal password is: XXXXX

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

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

Yours sincerely

Conversation guideline

Key date: 18.04.2017 2.05.2017 16.5.2017 30.05.2017 13.06.2017 27.06.2017

Store Manager: _____

What has the store manager done to increase the DB1?

What problems have occurred?

Which measures / which next steps does the store manager want to carry out until the next meeting?