Team incentives and performance: Evidence from a retail chain[1]

February 2015

Guido Friebel, Matthias Heinz, Miriam Krüger, Nick Zubanov

Abstract: We test the effectiveness of team incentives by running a natural field experiment in a retail chain of 193 shops and 1,300 employees. A bonus was offered to shop sales teams for sales surpassing the pre-existing targets. The team bonus, introduced in response to intensified market competition, is a natural choice for our study firm, where teamwork is required by technology and individual sales performance is unobservable. On average, team bonus increases sales and customer visits in the treated shops by around 3%, and wages by 2.3%. The bonus is highly profitable for the firm, generating an extra 3.8 $ of sales, and 2.1 $ of operational profit, per dollar spent. The treatment effect is heterogeneous along several dimensions, as predicted by our theoretical model, being larger in shops located in big towns, employing younger workers, and with fewer "mini-job" employees, who could not partake in the bonus program for tax reasons. Our results show the importance of complementarities within teams and suggest that improved operational efficiency is the main mechanism behind the treatment effect.

Keywords: management practices, randomized controlled trial (RCT), natural field experiment, team incentives, insider econometrics

JEL codes: D23, J33, M52

[1]All authors are at Goethe University Frankfurt, except for Heinz at University of Cologne. Friebel is also affiliated with CEPR and IZA, and Zubanov with IZA. We are grateful for the support of Deutsche Forschungsgemeinschaft (DFG). We would like to thank for their comments: Oriana Bandiera, Iwan Barankay, Nick Bloom, Thomas Dohmen, Florian Englmaier, Niels Kemper, Michael Kosfeld, Johan Lagerlöf, John List, Devesh Rustagi, Orie Shelef, Dirk Sliwka, Matthias Sutter and participants in seminars at Adelaide, Bergen, Birmingham, Bonn, Cologne, Columbia, Copenhagen, University of Otago, the EBRD, Frankfurt, King's College, Maastricht, Rotterdam, QUT in Brisbane, University of East Anglia, UNSW, a conference organized by the university of Arhus, a workshop organized by LMU Munich, the Annual GEABA conference and the NBER Organizational Economics Working Group Meeting 2014 in Stanford. We would also like to praise the team spirit of our partners in our study firm, and of Artur Anschukov, Sandra Fakiner, Larissa Fuchs, Andre Groeger, Daniel Herbold, Malte Heisel, Robin Kraft, Stefan Pasch, Jutta Preussler, Elsa Schmoock, Patrick Schneider, Sonja Stamness, Carolin Wegner, Sascha Wilhelm, Sandra Wuest, who provided excellent research assistance.

1. Introduction

"How can members of a team be rewarded and induced to work efficiently?" This question, asked by Alchian and Demsetz (1972) in their influential contribution to the economic analysis of organizations is at the heart of this paper. While Alchian and Demsetz argued for input monitoring by a manager-owner, incentives conditioned on joint output would be a natural alternative. However, because teamwork blurs the performance of individuals into a common performance signal, team incentives are weakened, except when contracts are feasible in which the team members are penalized unless the outcome reaches the efficient level (Holmström, 1982). These contracts are hard to implement in reality. Hence, in stark contrast to *individual* incentives that have been shown to work well in the field (Lazear, 2000; Shearer, 2004; Bandiera et al, 2009), it remains a rather open question whether team incentives are effective (Bloom and Van Reenen, 2011).

We bridge this gap by providing answers to the following related questions: (i) Do team incentives raise performance in the field, and by how much? (ii) What profits can be reaped, and what is the impact on wages? (iii) What conditions are conducive to team incentives' effectiveness and what can be learned from this heterogeneity of treatment effects about the anatomy of teamwork? (iv) Through what mechanisms do team incentives affect performance and to what extent are these in line with expectations?

Using a novel randomization method developed by Barrios (2014), we carried out a natural field experiment in a bakery chain in Germany in the time between April and June 2014. Shop sales teams consist of on average 7 people (4 full-time equivalents) who are hired and assigned to teams centrally; there is no movement between teams. In the randomly selected treatment shops, sales teams could benefit from the bonus we introduced. In line with the bonus schemes used in the firm for its managers on different levels, the sales team bonus was a step function, which, depending on sales above pre-determined targets, paid up to 300 € (at the time around 400 $) to the team, to be shared between the shop employees pro-rata based on the hours worked in the respective month. Sales targets had been agreed months before we started discussing the introduction of a team bonus, and the firm maintains a policy of non-renegotiation of targets. Because employees work in an ongoing firm and carry out their normal day-to-day job, our experiment combines realism and randomization (List and Rasul, 2011). Besides the change in the compensation scheme, there is no other intervention. We also took care to ensure that employees would not consider themselves part of an experiment. Except for our partners in management and the workers council, no one was aware of our involvement, and communicating the bonus scheme to the sales staff was taken care of by the management.

In these communications, the firm used the term "pilot", often employed by it when introducing new practices for a limited period of time. Finally, because of the firm's hierarchical HR processes, sorting of workers into different teams and incentive schemes is not an issue - unlike in Boning et al. (2007) and Hamilton et al. (2003) who established the importance of self-selection when teamwork is introduced. Our experiment thus looks into the effect of a change in compensation only.

The analysis of our data provides the following answers to the questions raised before. Team incentives increase sales by about 3%, equivalent to one third of the standard deviation. The effect is stable over the entire treatment period; it is also robust to changes in econometric specification and to a number of other checks, most importantly, contamination, Hawthorne effect, and gaming of the incentive scheme. Many of the shops in the treatment group increased their sales beyond the level at which the bonus was capped, indicating large potential efficiency gains associated with a simple bonus scheme. The bonus increased the wages of sales staff in the treatment group by 2.3 % (and up to 13% in some individual cases). The team bonus was highly profitable for the firm whose management decided to roll out the scheme to all shops in July 2014. Our calculations show that each dollar spent on bonus payment yields 3.8 $ extra sales, and 2.1 $ operating profit. This is a large effect for the retail sector in general, and particularly so for Germany.[2]

We propose a simple model whose predictions allow us to explore not only the overall treatment effect but also its heterogeneity to learn about the anatomy of teamwork. In our model, a team is incentivized by a bonus that is paid when sales are above a certain threshold. The first three predictions of the model are straightforward: incentivized teams put in more effort and sell more; team effort will increase with the marginal returns to effort, and decrease with the costs of effort. Indeed, there is an overall treatment effect of 3%, which goes up to 6% in shops in big towns, an increase of 2/3 of the sales' standard deviation. Arguably, this strong effect owes itself to higher demand in big towns, an explanation we discuss further below. The treatment effect is also more pronounced in sales teams with younger employees whose iso-utility curves in money and cost of effort are presumably flatter.

Turning to other, less straightforward predictions, we investigate how variations in N, the size of the team, should affect the magnitude of the treatment effect. Keeping the bonus

[2] According to Bloom and van Reenen (2007) and Bloom et al. (2012), Germany is one of the countries with the highest level of managerial efficiency in general, second only to the US. This also applies to retail (Bailey and Solow, 2001) a highly competitive sector, in particular because of the presence of two retail discounters, Aldi and Lidl, and low entry barriers (in contrast to, for instance, to France, see Bertrand and Kramarz, 2002). In fact it was precisely the entry of these firms into the market for fresh bread that triggered the change in incentives that we analyze here.

constant, one might expect that an increase in *N* would reduce the treatment effect, because each member would receive a smaller bonus. However, our model shows that increasing *N* can reduce or increase the treatment effect depending on the strength of complementarities and the curvature of the cost of effort function.[3] We can use an institutional specificity in Germany as a source for exogenous variation of *N*, reinterpreted as the share of non-incentivized members of a team: roughly a third of the workers in our shops are so-called mini-jobbers who earn up to 450 € per month tax-free. For tax reasons, these employees were not eligible for the bonus. The share of hours worked by these employees in the shops is orthogonal to the treatment, providing identification for the effect of *N* on effort given the bonus. We find that the treatment effect drops rapidly in shops with a higher proportion of mini-job workers, confirming the importance of complementarities within teams. A final prediction of our model is that shops with a worse performance record will react more strongly to the bonus, which is also confirmed.[4]

As to the fourth research question - on the mechanism - our results show that the treatment effect is unrelated to upselling (higher sales per customer visit). To increase the sales per customer was an important element of the firm's strategy and reflected in sales guidelines and managerial activities. Both the management and we expected that a treatment effect should mainly manifest itself in this channel. It appears instead that the team bonus provided a stimulus for improving operational efficiency. This interpretation is supported by three findings. First, the treatment effect on sales and customer visits are of a similar magnitude; second, the treatment effect is largely driven by shops in big towns, where more efficient operations can increase demand by cutting queue lengths and waiting time; third, we find no evidence for alternative explanations, such as management input or "working smarter" (Burgess et al., 2010) through work shift reallocation. We hence believe that the treatment effects cause is sales assistants' increasing efficiency in carrying out their tasks under unchanged organizational arrangements.

Before discussing the implications of our experiment, we would like to stress that the type of work in our partner company is teamwork par excellence, because it combines potential complementarities between team members' efforts and unobservability of individual performance signals. Notice that both are ingredients of Alchian and Demsetz's (1972) theory,

---

[3] This result reflects some of the insights from the political economy literature on the "group-size paradox" (Esteban and Ray, 2001).

[4] This finding demonstrates the importance of the design of a compensation scheme for determining effort choices by heterogeneous agents. Our scheme, being non-competitive, elicits greater response from relatively unproductive teams. On the other hand, the tournament-based incentives in Delfgaauw et al.'s (2014) experiment induced historically best performing shops to put in most effort.

while complementarity is not a necessary ingredient to Holmström's (1982) theory. "Our" chain always operated in teamwork organization in which each worker carried out a broad variety of tasks, including handling the goods delivered, preparing food in the oven, taking care of the customers, and handling the cash register. The time workers spend on each task varies much, employees work in overlapping shifts and they are supposed to help each other. The need to deal with different tasks that occur at unknown frequency makes it too costly to have highly specialized agents who would be idle most of their time. Providing individual incentives in such a setting would lead to measurement and potentially gaming problems and productivity losses because employees may cut help efforts for other members of the team (Itoh, 1991, Auriol et al, 2002). All of these elements distinguish our setting from the ones in Boning et al (2007) and Hamilton et al (2003) in which the intervention is to change from individually-based to team-based work organization *and* compensation, while in our setting there is no discretion in the design of the workplace, such that the relevant question is confined to the effect of performance-related team compensation.

Because of the widespread use of teamwork that is similar to the one in our study firm, we believe our results to be widely applicable. The study context is representative of many firms and jobs in the global economy. Indeed, retail is one of the largest industries in the world in terms of employment. In Germany, more than 3 million people (7% of the labor force) work in retail. Not only in retail, but also in many other service industries, for instance, restaurants, hotels, airlines, similar workplace organization is present: employees carry out many tasks, are supposed to help each other, and their individual performance is difficult or impossible to measure.

In addition to the main result of our study – team incentives work – several implications of interest for researchers and practitioners alike follow from our findings. In particular, treatment effect heterogeneity may be taken as guidance for maximizing the effect of team incentives in firms similar to ours. A profit-maximizing firm would thus do well by promoting team incentives in more urban and currently underperforming shops with younger, equally incentivized workforce – provided the information spillovers between shops have no negative consequences for motivation. We could not detect any spillovers.

While treating different shops differently may be profit maximizing, unequal treatment *within* teams is detrimental for performance because of effort complementarities within teams. Indeed, the team bonus effect falls quickly with the share of unincentivized mini-jobbers. Peer pressure (Kandel and Lazear, 1992, Mas and Moretti, 2009) thus seems to have limits, as it does not prevent noticeable performance losses from having more than 10% of mini-jobbers' work in full-time equivalent. Mini-jobbers are, however, not fully representative of the workforce as they tend to have high turnover and low identification with their teams, and may hence not (much) react to peer pressure in the first place.

Our paper contributes to the existing empirical literature on incentives by providing clean causal evidence on the performance effects of team incentives in the field,[5] free from nonrandom treatment assignment and worker self-selection into teams. The technology, strategy and basic management practices are common across the teams in our sample, thus avoiding unwanted heterogeneity. The uniformity of context and randomization combined with realism make our natural field experiment different from previous studies in which the adoption of teamwork or team incentives is arguably nonrandom. Boning et al. (2007), Hamilton et al. (2003), Bandiera et al. (2009) and Bandiera et al. (2013) all find supportive evidence that teamwork and incentives raise efficiency. However, the first study shows that the decision in favor of teamwork and its effects depend on technology, while the other studies observe nonrandom sorting of workers into teams. In particular, Hamilton et al. (2003) find that more productive workers prefer teams so strongly that they even forgo individual earnings in exchange for working in a team. The questions and context are different to ours, since in our study teamwork is technologically fixed and only compensation schemes vary. Our work is also different from Delfgaauw et al. (2013) who are interested in testing tournament theory in the field.

Another contribution of our work is to provide an insider perspective on the adoption of management practices (Ichniowski et al, 1997), complementing the literature on management practices across firms. Our study confirms the interpretation of Bloom and van Reenen (2010), Syverson (2011) and Bloom et al. (2014) who argue that one of the main reasons why some firms adopt productivity-enhancing management practices and others do not, is product market competition. Additionally, building on our insider knowledge, we highlight the importance of internal firm politics as another factor influencing the adoption and success of new management practices. In particular, we argue that the same institutions that in some instances create inertia and resistance to change, such as worker councils in Germany, may, in other instances, be conducive to reach Pareto improvements between management and workers, because they are able to create high levels of commitment.

2. Background

[5] Nalbantian and Schott (1997) have investigated these incentives in the lab.

2.1 Changes in the market and the challenges faced by the study firm

In the period between the 1980s and early 2000s, German bakery chains like ours, some of them owning hundreds of shops, had successfully built their business model exploiting the benefits of attractive locations, such as supermarkets and malls, and economies of scale. The chains had crowded out many of the existing small master bakeries whose number and market shares had steadily declined. In 2011, however, discounter retailers *Aldi* and *Lidl* began to sell freshly baked bread and related products in their dense network of existing shops, with large success. Their bread is widely believed to be of similar quality as the one in the bakery chains, but is sold at much lower prices, thus forcing the incumbent chains to rethink their business model.

As a consequence, many of the chains, including our study firm, started differentiating their product range, moving into the market for snacks, cakes, sandwiches and beverages traditionally covered by cafés and fast food chains. This strategic move was accompanied by substantial investments in shop design to make it more attractive and inviting. Prices for different kinds of products were adjusted, and additional marketing instruments were introduced, such as special weekly offers, sales related to charitable activities (for each bread bought, the firm donates x cents to a local charity). Furthermore, the HR practices were put under scrutiny. In the past, true to the English saying "something sells like hotcakes" and its German equivalent "something sells like sliced bread", many employees of the firm had taken the steady demand for granted and many members of the middle management had failed to motivate their subordinates in the shops to actively engage with the customers. However, with the changes in the market situation the firm reacted to the challenges, and tried to develop new HR management practices aimed to improve shopping experience.

2.2 The firm's HR management practices before the market change

Before the changes in the market discussed above, our study firm operated different sets of practices for managers and sales staff. For managers at all levels of hierarchy - shop, district and top - there was a detailed system of key performance indicators (KPI) according to which they were evaluated and paid. For top managers, the KPIs consist of sales, profit, and strategic outcomes, for example, sales of a certain product. For district managers, who oversee 10 to 15 shops in a certain area, the KPIs consist of sales, personnel costs and customer service evaluations obtained from monthly mystery shopper visits in their area. For shop managers, the KPIs are the same as for district managers, except that they are based on the performance of their shops alone. Sales is by far the most important KPI for managers at all levels in terms of their bonus. Sales performance is incentivized by offering a bonus which depends on reaching a sales target determined in the end of the preceding year based on past sales and correction for the general trend in sales (-2% in 2014).

Unlike managers, sales agents, who make up about 80% of the staff, received fixed wages only (9-11 Euros per hour, depending mainly on tenure). The fixed wages for all sales agents are determined by collective agreements. There are two groups of sales agents: regular ones, whose income depends on their hours and who pay regular income tax, and the so-called “mini-jobbers”, workers who, in addition to receiving welfare benefits, earn up to 450 Euros per month tax-free. The sales agents are predominantly unskilled, and employee turnover is high (see descriptive statistics in section 3.3), making the profitability of investments in training questionable. Instead, the firm traditionally operated only a limited set of HR practices applicable to its sales staff, relying on shop manager supervision to ensure compliance with operational procedures (serving customers, handling goods, etc.).

Given the recent market changes, and the firm's strategic response to them, the HR practices offered to sales staff were no longer perceived as optimal by the top management. After experimenting, unsuccessfully, with hiring more qualified employees to improve customer service, the firm approached us in 2013 for advice on a feasible set of HR practices to support their new strategy. We agreed to help provided we would have access to all data required, and would be free to test the effectiveness of a new HR practice according to the requirements of a “natural field experiment” (Harrison and List, 2004). We received sales, financial and accounting, geographical, compensation and personnel data of the shops since January 2012, allowing us to carry out a very precise randomization procedure, which is explained in more detail in section 3.3. We offered our advice free of charge and covered most of the research costs. The company committed itself to providing the data and all administrative support needed. Our main interfaces were the CEO of the company, the Head of HR and his team, the Head of Sales, and a small selected group of district managers.

2.3 Proposed changes

Given the substantial number of HR and other practices the company had experimented with before our involvement, and the existing well-functioning system of performance measurement, in particular, concerning sales, we (the researchers) converged quickly on the idea of implementing a team bonus, leaving unchanged all existing practices. In late February 2014, we then proposed to our firm introducing a bonus payable to shop sales teams conditional on reaching or exceeding the sales targets already in place for managerial bonuses. The firm's first reaction to our proposal can be summarized by the response of a surprised member of the management team: “[Monetary incentives to sales staff] were simply never on our agenda.” Other members of the management team were afraid that bonus payments could become a burden on the firm that already had its profit margins reduced by intensified competition. Indeed, in addition to the payments to shop teams reaching their sales targets, there would be a knock-on effect on the bonuses paid to managers given the existing incentives, a sizable effect as we discuss later.

In response to these concerns, we ran simulations of the bonus' effects on sales and personnel costs. Our simulations showed that the expected team bonus payments would be lower than 20,000 Euro per month in case half of the shops were treated and the maximum monthly bonus was capped at 300 Euros. This convinced the top managers to try a “pilot” study with half of the shops assigned to the team bonus scheme. However, the district managers in the task force were afraid that the subsequent rise in the wage costs would reduce their own bonus if the wage costs targets remained fixed. Top management then decided that the bonus payments to sales staff would be made from a different budget and would not affect the personnel costs relevant for district managers’ KPIs. The district managers were quick to realize that in such a setting they were likely to benefit as well if the team bonus increased sales in the shops under their supervision. The worker council also was in favor of the bonus, in particular, because it was designed as a pure add-on payment and thanks to the high level of trust between the council and management. As we will argue later, this coalition with the worker council may have been crucial for being able to carry out the experiment.

3. Experimental procedures

3.1 Preparation

We began our preparations for the experiment by planning two waves of an employee survey, the first in March 2014, a month prior to the introduction of the team bonus, and the second in May, in the middle of the treatment period. We conducted the surveys primarily for three reasons. First, to see whether there is a treatment effect on employee attitudes – an issue deemed important by the management; second, to check whether our treatment and control samples are balanced with respect to employee attitudes; third, to test whether baseline attitudes affect the response to our treatment.

The main variables we measured in both waves of the survey using the metrics developed by Allen and Meyer (1990) were employee attitudes indices: satisfaction with the job context, overall satisfaction, and organizational commitment. The May survey collected some additional data we used for robustness checks (more details in sections 6). The surveys were distributed through the district managers and collected by our research assistants in sealed envelopes as an extra guarantee of anonymity. Our logistics and communication efforts helped secure response rates of 80 percent in the first and 60 percent in the second wave of the survey. We have found no treatment effect on either dimension of employee attitudes, nor any significant interaction between baseline attitudes and the treatment effect on sales. Therefore, to save space, we will concentrate on the treatment effect on sales and customer visits in what follows.

In preparation for the team bonus, we designed information leaflets to be placed in the back offices of the treatment shops, and letters to be distributed by the district managers to the employees. In contrast to the employee survey, the logo of Goethe University did not show on these materials (see Appendix I), so that people would not perceive themselves as part of an experiment. In fact, there was no mention of our research team in any communication regarding the bonus. Apart from top management, the only group of employees who knew the allocation of shops into the treatment and control groups were the district managers. In a meeting on March 25th 2014, we told all of them about our team bonus experiment for the first time and handed to every manager the list of the control and treatment shops in their district.

We trained district managers at the same meeting in how to explain the team bonus to the shop managers in the treatment group who would then relay our explanation to the employees in their shops. We also instructed the district managers in how to react to questions about the bonus from the employees in the control group shops. Should these questions be asked, the district manager would respond: "This is a pilot. Every shop had the same chance to be drawn into the bonus scheme. The work council agreed to this procedure." It was the worker council who suggested that this response would be acceptable for the employees in control shops in case they found out about the bonus scheme. We called the district managers every second week to enquire whether employees in the control group had heard about the team bonus. It turned out that questions about the team bonus were seldom asked. In section 6.1 (contamination) we discuss this and other procedures to detect contamination in more detail and show that there was no evidence for contamination.

We also explained to the district managers, as well as wrote in the information leaflets sent to the treatment shops, that mini-jobbers had to be excluded from the bonus scheme due to tax reasons. Namely, if a mini-jobber earned more than 450 Euros a month she or he would end up paying taxes on their *entire* income. Therefore, giving bonus to mini-jobbers would reduce, rather than increase, their net wage. According to the district managers we interviewed, the mini-jobbers accepted this and no complaints were raised.

3.2 The bonus scheme

Figure 1 illustrates the bonus scheme offered to the treatment shops. Shops that reach the sales target for the month receive a bonus of 100 Euro to be shared between the part-time and full-time employees in the shop in proportion to their hours. The bonus increases by 50 Euro for each percent point above the target and is capped at 300 Euro per month for exceeding the target by 4 percent or more. Hence, the team in a shop can make extra earnings of up to 900 Euro in the treatment period of April to June 2014. We provided the employees with examples of what sales increases would mean in terms of additional goods to be sold per day (for instance a one-percent increase above the sales target for a mid-sized shop would be tantamount to selling per day ten additional rolls, two loafs of bread, some sandwiches and some cups of coffee).

FIGURE 1 ABOUT HERE

We realize that this bonus scheme may be criticized on theoretical grounds for being susceptible to strategic behavior of employees around the bonus cutoffs (we label this as "gaming" in what follows). However, designing an incentive scheme one always faces a tradeoff between optimality on one hand, and clarity, verifiability and approval of the scheme by its stakeholders on the other. Our bonus scheme reflects this tradeoff, which in fact is not too specific to our study environment since "step-wise" bonus schemes like ours are rather widely spread.[6] We do nevertheless address the possibility of gaming in section 6.3.

3.3 Randomization and power of the experiment

We follow Barrios (2014) who shows that randomizing pairwise by using the predicted outcome variable, in our case sales, minimizes the variance of the difference-in-difference treatment effect estimates. We use historic observations between January 2012 and December 2013 to run a regression of log sales on labor input with month and shop fixed effects, from which we obtain predicted sales. We then rank the shops according to the predicted sales and randomize within the pairs of shops with adjacent ranks, except for the median-ranked shop (#97) which we randomly assigned to the treatment group. The resulting treatment and control groups comprised 97 and 96 shops, respectively. The sample size is sufficient: power calculations on the basis of 27 months of observations pre treatment (January 2012 to March 2014) and 3

[6] For example, the *World at Work* 2012 survey of incentive pay practices in ca. 200 large U.S. private firms finds that some form of incentive pay is practiced in 95% of the sample. Of the firms that do practice incentive pay, 88% offer performance bonuses.

months post treatment (April to June 2014) show that we would need 70 shops in each group to detect a 3% treatment effect at a 5% significance level with the probability 0.9.

TABLE 1 ABOUT HERE

Table 1 summarizes their pre-treatment characteristics. Thanks to our randomization procedure, the treatment and control samples are balanced in the average pre-treatment sales, our key outcome variable. They are also similar in other potentially relevant characteristics, such as the percentage of unsold goods, number of customer-visits, frequency of achieving the sales target, location, and employee attitudes. In fact, none of the averages reported in Table 1 differ significantly between the groups. An average shop sells over 27,000 Euros worth of goods[7], employs 7 people most of whom are female in their late 30s, unskilled, and working part-time. There is a sizeable share of workers on a mini-job, around 30%, who for tax reasons were excluded from the team bonus scheme. Sales are quite variable, with location and size differences explaining 90% of the variance. There is also a considerable variation within shops, much of which is due to seasonal demand, temporary closures for renovation, and market dynamics, such as the entry and exit of competitors, all of which factors we control in our statistical analysis.

FIGURE 2 ABOUT HERE

Figure 2 displays spatial distribution of our control and treatment shops. The region in which our partner firm operates spans roughly 100 km from West to East and 60 km from North to South, an economy of more than 3 million inhabitants. Shop locations vary in population size. However, almost all shops are placed on the premises of supermarkets owned by the parent company, or in their immediate vicinity, relying thus mostly on the customer traffic to and from grocery shopping.

4. Model

An illustrative model is helpful to generate hypotheses about the average treatment effect, and potential treatment heterogeneity.

4.1 Basic setup

[7] One shop, located at a local transportation hub and assigned randomly to the treatment group, sold on average 118,000 euros worth of goods per month in the pre-treatment period and employed 22 people. Excluding this shop, the average pre-treatment sales in the treatment group are 27,176 euros per month with standard deviation of 10,885 euros, which is much closer to the same characteristics of the control group. Removing this shop from our regression sample does not change the estimated treatment effects.

Consider N agents working in a team creating output y that depends on total effort E, a parameter measuring the productivity of team effort, a, and noise v with a probability distribution function $\phi(v)$ symmetric around, and centered at, zero:

$$y = a \cdot E + v \tag{1}$$

Total effort is assumed to be a CES aggregate of individual efforts $e_i, i = 1, \dots, N$:

$$E(e_1, \dots, e_N) = \left(\sum_{i=1}^{N} e_i^p \right)^{\frac{1}{p}} \tag{2}$$

The effort aggregation in (2) is flexible and can accommodate effort complementarity (p < 1) or substitutability (p > 1); when p = 1, the team's total effort is the sum of individual efforts.

To model the incentive effects of the bonus scheme used in our study firm, we consider that a team bonus $B > 0$ is paid if and only if the output exceeds a performance target y_0. To keep the complexity of the model to a minimum, we only consider one such performance target rather than the multi-step bonus scheme that we implemented (Figure 1). The expected bonus is

$$g(E) = B * \text{prob}(a \cdot E + v \geq y_0) + 0 * \text{prob}(a \cdot E + v < y_0) = B\Phi(a \cdot E - y_0), \tag{3}$$

where $\Phi(a \cdot E - y_0) = \int_{-\infty}^{a*E-y_0} \phi(v)dv$ is the cumulative density function of v.

The bonus is split evenly between team members who individually decide on the level of effort e_i to contribute towards reaching the team performance target by maximizing their own payoff:

$$\pi(e_i, e_{-i}) = \frac{1}{N} B\Phi(a \cdot E - y_0) - b \cdot c(e_i), \tag{4}$$

where $c(e_i)$ is the costs of effort function, assumed to be continuous, twice-differentiable and convex, and b is a parameter measuring the difficulty of effort. The effort choice is constrained from below by a "minimally acceptable level" e_0, which could be owing to some intrinsic motivation as in Holmström and Milgrom (1991, p. 33) or to some monitoring activity by the firm as in Lazear (2000). There is also a maximum possible level e_{max}, and both levels are assumed to be the same for all team members.

Assume for the time being that the parameters of the payoff function (4) are the same for all team members (we will introduce heterogeneity in the payoff function later). With complementarity, there is a continuum of symmetric equilibria. We focus on the equilibrium in which members choose the same optimal effort $e_0 \leq e^* \leq e_{max}$ satisfying any one of the following sets of conditions:

$$\left.\frac{d\pi}{de_i}\right|_{e_i=e^*} = aN^{\frac{1-2p}{p}} B\Phi'\left(aN^{\frac{1}{p}}e^* - y_0\right) - b \cdot c'(e^*) \quad = 0 \tag{5}$$

$$\left.\frac{d\pi}{de_i}\right|_{e_i=e_0} > 0$$

$$\left.\frac{d^2\pi}{de_i^2{}_i}\right|_{e_i=e^*} = N^{\frac{2-2p}{p}} Ba^2\Phi''\left(aN^{\frac{1}{p}}e^* - y_0\right) - b \cdot c''(e^*) < 0$$

or

$$e^* = e_0 \text{ and } \left.\frac{d\pi}{de_i}\right|_{e_i=e_0} \leq 0, \tag{6}$$

or

$$e^* = e_{max} \text{ and } \left.\frac{d\pi}{de_i}\right|_{e_i=e_{max}} \geq 0$$

In words: there will be interior solution given by (5) if the marginal benefit of effort exceeds its marginal costs at the minimum acceptable level e_0 but is below the costs at the maximum possible level e_{max}, and if the payoff function $\pi(\cdot)$ is concave in effort.

4.2 Predictions

We derive the following testable predictions for the interior solution $e_0 < e^* < e_{max}$ from the comparative statics on the first-order conditions (5):

1. The effect of the bonus on individual effort, and hence expected output, is positive:

$$\frac{de^*}{dB} = -\frac{aN^{\frac{1-2p}{p}}\Phi'\left(aN^{\frac{1}{p}}e^* - y_0\right)}{\frac{d^2\pi}{de_i^2}} > 0$$

2. Individual effort increases with the productivity parameter a (under reasonable assumptions):

$$\frac{de^*}{da} = -BN^{\frac{1-2p}{p}} * \frac{\Phi'\left(aN^{\frac{1}{p}}e^* - y_0\right) + N^{\frac{1}{p}}e \cdot a\,\Phi''\left(aN^{\frac{1}{p}}e^* - y_0\right)}{\frac{d^2\pi}{de_i^2}} > 0.$$

The expression in the numerator of the derivative of $e^*$with respect to a,is positive assuming $\left|\Phi''\left(aN^{\frac{1}{p}}e^* - y_0\right)\right| \ll \Phi'(aN^{\frac{1}{p}}e^* - y_0)$, which is the case when the team's output, $aN^{\frac{1}{p}}e^*$, is close to performance target y_0 and N is not too large. [8]

3. Individual effort decreases with team size N if effort complementarities are not too strong ($p \gg 1/2$). However, depending on the strength of effort complementarities and the convexity of the costs of effort function, the team's *total* effort may increase or decrease with N. See Appendix II for the proof.

4. Individual effort decreases with the difficulty of the costs of effort parameter b, since

$$\frac{de^*}{db} = -\frac{-b}{\frac{d^2\pi}{de_i^2}} < 0$$

5. Team effort decreases with the share of non-incentivized members in the team. See Appendix II for the proof.

6. The effort under the bonus will depend on the frequency of reaching the targets in the past, without the bonus. More successful teams' effort response to the bonus will be weaker than that of less successful teams. However, depending on the costs of effort, extremely unsuccessful teams may not respond to the bonus at all, choosing the corner solution $e^* = e_0$ instead. See Appendix II for the proof.

5. Results

5.1 The effect of team bonus on sales

We now turn to our basic result, corresponding to prediction 1 of our model, that the introduction of team bonus increases output through higher effort exerted by the now incentivized workers. Table 2 reports the treatment and control shops characteristics in the treatment period (April to June 2014), giving a first impression of the treatment effect. Sales and customer-visits have gone down, reflecting the secular downward trend in the bakery business. Yet, the drop in sales and customer-visits being less pronounced in the treatment than in the control group suggests a positive treatment effect. In fact, the difference-in-difference

[8] Note that $|\Phi''(x)|/\Phi'(x) = x$ for the standard normal distribution, less than x for fatter-tailed distributions, and 0 for the uniform distribution. So, our assertion that $|\Phi''(x)| \ll \Phi'(x)$ for x close to zero is true for many distributions.

estimated effects on the log sales and customer-visits are 3.3% and 2.8%, respectively, both significant at conventional levels. Since there is no significant treatment effect on other outcomes, we proceed with a more in-depth analysis of sales (this section) and customer-visits (section 7.1).

TABLE 2 ABOUT HERE

To visualize the treatment effect on sales, Figure 3 plots the treatment and control groups' year-on-year sales growth in the treatment month versus the sales levels in the same months (April to June) of 2013. Additionally, Figure 4 displays the kernel density graphs of the year-on-year sales growth for the two groups. The shift in the treatment group's sales growth distribution to the right from the control group's is fairly uniform across the growth rates and initial sales levels.

FIGURE 3 AND 4 ABOUT HERE

We estimate the treatment effect from the following baseline difference-in-difference specification:

$$\ln(sales_{it}) = \beta * treatment_i * after_t + period_t + shop\ fixed\ effect_i + controls_{it} + error_{it} \qquad (1)$$

where $\ln(sales_{it})$ is the log sales in shop i and month t, the *treatment* dummy takes the values 1 for the treatment and 0 for the control group shops, the *after* dummy is 0 for the periods before treatment and 1 thereafter, $controls_{it}$ include the log total hours worked and dummies for renovation within the last two months, and $error_{it}$ is the idiosyncratic error term which we cluster at the shop level to allow for serial correlation. (Bootstrapping produces standard errors of similar magnitude.) Coefficient β is the difference-in-difference estimate of the average treatment effect, measuring the percentage increase in sales caused by our treatment.

The estimates based on our baseline specification (1) are presented in Table 3 with (column 1) and without (column 2) “outliers” defined as observations with year-on-year sales growth exceeding 30% in magnitude. The average treatment effect is an upwards of 3% and is statistically significant.

In addition to clustering errors at the shop level, which may still underestimate coefficient standard errors in small samples (Cameron and Miller, 2015), we implement another solution, originally proposed in Bertrand et al. (2004) – to estimate our baseline specification with only two observations per shop, one pre- and the other post-treatment average (column 3). As another robustness check to our baseline results, we allow for the correlation between the treatment status and the baseline outcome, which, despite randomization, may occur in finite samples and causing the “regression towards the mean” problem (Stigler, 1997). Specifically, we introduce two modifications. First, we augment the two-period specification discussed above with the log average sales before treatment (column 4). Second, we run our baseline specification with sales growth relative to a specified base as the dependent variable, including the base sales as control (columns 5-7).

TABLE 3 ABOUT HERE

Whatever specification we use, we obtain the average treatment effect estimates of similar magnitude – around 3% – and significance. This uniformity suggests that neither of the estimation issues we mentioned above and addressed in our analysis is important in our data. Indeed, simply clustering the errors by shop is sufficient on the relatively large sample such as ours. Regression to the mean is not a concern either since our sample is well balanced. Calculating the treatment effect in each month with our baseline specification (1) as an extra robustness check (see Table 4, Panel A, column 1), we find it to be 2.9% in April 2014, 3.7% in May, and 2.9% in June 2014, a steady effect without noticeable abatement.

Let us gauge the profitability of our team bonus scheme by comparing its implied gains with the total economic costs of its implementation. The estimated average treatment effect on sales of 3% implies an extra 820 Euros (=[exp(0.03)-1]*27,000) worth of sales per month in the average shop, or 238,620 Euros (=820*3 months * 97 shops) in all treatment shops over the treatment period. Given the historic share of value added in sales at 0.56, the implied operational profit gain is 133,627 Euros.

Turning to the costs, around 50% of the workers in the treatment group received a bonus at least once in the treatment period. The total bonus averaged at 114.7 Euros or 3.9% of the average recipient's quarterly earnings. The total bonus payments made by the company in April to June 2014 amounted to 35,150 Euros, or 2.3% of the total labor costs in the treatment shops. There was a knock-on effect on shop manager bonus: 240 Euros per treatment shop per quarter (= difference-in-difference estimate of the treatment effect on shop manager's bonus), adding an extra 23,280 Euros for all 97 treatment shops. In addition, there is an estimated effect on the district and top manager bonus of 4,500 Euros. There were also one-off costs of activities necessary for the implementation of the bonus scheme: printing and delivering posters and other materials, administrative support (bonus calculations and communications) and the costs of managers' and researchers' time required to implement the scheme, which we estimate at 25,000 Euros. [9] The total costs add up to 87,930 Euros.

[9] This estimate excludes the costs of research activities not directly related to the bonus, such as surveys and mystery shopping.

The implied benefit from the scheme net of the costs is 45,700 Euros for the treatment period and for the treatment shops. However, projecting our calculations to the time past July 2014, when the scheme was rolled out to all 193 shops and little overheads were still necessary, the implied net gain becomes 140,000 Euros per quarter for the entire chain. Overheads aside, our calculations imply that each dollar spent on the bonus brings 3.8 dollars of extra sales, or 2.1 dollars of extra operational profit. In sum, our scheme is a viable "investment in people" project and a win-win for the firm and its workers.[10]

5.2 Treatment effect heterogeneity

Ichniowski and Shaw (2012) argue that the effect of a new management practice often differs between workers and workplaces even under the same production technology, encouraging researchers to "estimate the production function with heterogeneity in the management treatment effect" (p. 265). Indeed, our model predicts heterogeneities in the effect of team bonus along several dimensions: shop location, shop workforce size and composition, and success in reaching the sales target in the past. In what follows, we report the results of testing these predictions collected in Table 4. It is worth repeating (see Table 1) that our treatment and control groups are balanced in all the characteristics we analyze below.

TABLE 4 ABOUT HERE

5.2.1. Shop location (Prediction 2)

Shop location affects the magnitude of effort's response to a given incentive by influencing the marginal product of effort. Thus, extra effort pays more in populous, urban locations that have office workers who might come in for lunch, and visitors who might buy a snack on the go; incentivized sales agent might cater to both these groups by improving operational efficiency (thus saving their time) and/or shopping experience. On the other hand, smaller locations have mostly regular shoppers whose demand for bread is harder to affect - hence the lower marginal product of sales effort in those locations. Besides, shops in urban locations have more competitors nearby, whose customers may be won over.

[10] Another project the firm undertook was to invest in a thematic redesign of thirty-one selected shops. However, the profitability of this project seems to be far less than that of the bonus scheme. Estimating the sales response in up to ten months after a shop was redesigned, we find the long-run average effect of 10% per month (probably an overestimate because of nonrandom selection). With the costs of redesign of at least 150,000 Euros per shop, the historic share of value added in output of 0.56, the German corporate tax rate of 30% (needed to calculate tax rebate), and a liberal lending interest rate of 3% per year, an average redesign project's return on investment over a ten-year horizon would be a mere 0.6% a year.

Table 4's Panel A reports the treatment effect by shop location. As expected, the treatment effect is largest, at 6%, in shops located in big towns (>60,000 inhabitants), going down to 3.8% in midsize towns, and zero in villages. As before, the treatment effect is fairly stable in time.

5.2.2. Workforce size and composition (Predictions 3, 4 and 5)

Shop workforce size will influence the magnitude of the treatment effect by increasing the total effort as the sum of individual efforts, as well as by decreasing the individual effort through free-riding. As we demonstrated in our model, which of these two opposite tendencies will prevail depends on the team production technology and the individual costs of effort function. To capture the variation in the treatment effect with workforce size, we interact the treatment dummy with the dummies for the quartiles of the shop-average number of workers not on a mini-job, thus allowing for nonlinearities in the treatment effect by size. Table 4's Panel B shows that the treatment effect is larger in bigger shops, and that the observed differences in the treatment effect do not owe themselves to bigger shops being located in bigger towns (column 2). In fact, the treatment effect increases with shop size in big towns faster than elsewhere.

Turning to the shop workforce composition, we explore treatment effect heterogeneity with shop workforce age, and the share of mini-job workers. We expect the treatment effect to be larger for younger workforce, since younger workers might have lower effort costs. Besides, there may be an element of resistance to change, which is weaker among younger workers, in the individual responses to our novel treatment. Table 4's Panel C reports treatment effects in the shops below and above the median workforce age on the whole sample as well as separately in big towns and elsewhere. Consistently with our expectations, “younger” shops respond to treatment more strongly. A further analysis suggests that the differential response to treatment by age is not driven by tenure: running our preferred difference-in-difference specification (1) with the treatment effect interacted with age and tenure separately as well as jointly produces a significant interaction with age but not with tenure.

The higher share of mini-jobbers should decrease the response to treatment, reflecting the drop in the size of the incentivized team. There will also be an additional negative influence if there are effort complementarities between mini-job and ordinary workers, since stronger complementarities increase the weight of the least productive workers' contribution to their team's output.[11] To accommodate the later, nonlinear, effect, we rerun our regression specification with the treatment dummy interacted with the dummies for each quartile of the shop-average share of mini-job workers, reporting the results in Panel D of Table 4. We find that the treatment effect goes down with the share of mini-job employees, especially in the shops located outside big towns. The abrupt drop in the treatment effect to zero past the second (whole sample) or first (shops outside big towns) quartile of the average mini-job worker share implies a steeper than linear decrease, which suggests effort complementarities between mini-job and ordinary workers in shop teams.

5.2.3. Past sales target achievement (Prediction 6)

We expect the treatment effect to vary with the past performance around the sales target. Historic record of achieving sales targets is informative for shop teams to gauge their probability of success in the future, since the targets are largely based on past sales (with a correction for the overall trend, hence the higher frequency of reaching the target in both groups, recall Table 2) and set in the beginning of the year. Our model predicts that less successful shops will respond to incentives more strongly - unless their past record is so weak that the prospects of reaching the target are not worth exerting effort above the minimum acceptable.

The last two panels of Table 4 report treatment effect estimates by quartile of historic distance to the sales target measured as: i) the difference between actual and target sales averaged for each shop over the pre-treatment period (Panel E1); and ii) the frequency of a shop achieving its target in the pre-treatment period (Panel E2). Shops in the bottom three quartiles of the distance to the target reacted to the treatment more strongly than did those in the top quartile, suggesting that rewarding the attainment of too easily achievable targets is not an effective motivator, and that team incentives can improve the performance even in quite unsuccessful shops.

6. Robustness checks

We consider three possibilities through which our estimated treatment effect may in fact have been the result of factors unaccounted for in our theoretical model. First, the treatment shops may have affected sales in nearby control shops through either carving into their sales or control shops workers sulking upon hearing that they were not part of the bonus scheme (we labeled

[11] As an example of the empirical framework required here, Iranzo et al. (2008) estimate a constant elasticity of substitution production function of different workers' skills within their firms. They find skill complementarity between, and substitutability within, occupational groups.

this possibility as contamination in section 6.1). Second, workers in the treatment shops may have worked harder than their immediate utility maximization would have them do, in order to try to increase the chance of the bonus scheme to be continued (a variant of the Hawthorne effect). Third, workers in the treatment shops may have "gamed" the bonus system by calibrating their sales effort so as to just meet the bonus target without going much beyond it (as documented in Courty and Marschke, 1997). In what follows, we discuss these possibilities and argue that they could not have explained our results.

6.1. Contamination

We have taken great effort to prevent contamination between the treated and non-treated shops in our experimental design. We did not let the workers in the control group know there was a team bonus in some other shops (the treatment group did not know there was a control group, either). We also developed communication protocols for the district managers to handle information spillovers between treatment and control shops so as to emphasize the fairness of the treatment assignment procedure. Additional measures we implemented to detect contamination during the experiment were: questions about inter-shop employee contacts in the second wave of the employee survey, bi-weekly communications with district managers, monitoring the firm's Facebook page, and controlling for the number of control bakeries in the neighborhood of a treatment bakery, and vice versa.

When we asked employees in the second wave of the survey about their contacts with colleagues in their and other shops during the treatment period, 83% of the respondents indicated that they had never mentioned the team bonus talking with employees from other shops. There is not much inter-shop communication in general: 80% of the respondents never spoke to a colleague from another shop. Consistent with finding little communication between employees, we learned from the district managers that only two employees from two control group shops asked them about the bonus, both in April. They received answers according to our protocol, which they found to their satisfaction. Removing the shops where bonus communications were detected or possible given the questionnaire answers did not change the baseline result. Finally, we inspected the firm's Facebook page, which attracts employees and customers alike who (sometimes to the dissatisfaction of the management) discuss internal issues such as stress at the workplace, quality of products, or problems of leadership and organizational culture. We could not find a single entry on the team bonus.

Finally, turning to the number of shops in the neighborhood as a proxy for the possibility of contamination, we interact the treatment effect with the number of other-group shops within a one-kilometer radius. This is the radius within which both contamination effects - business stealing and employee sulking - may reasonably be expected to occur. The treatment effect in this specification is 2.8%, close to the baseline, and interaction coefficient is insignificant (p-value=0.5). Summarizing, all our contamination tests fail to give evidence for contamination.

6.2 Hawthorne effect

There are several arguments against interpreting our results as a manifestation of the Hawthorne effect. First, as in Bloom et al.'s teleworking study (forthcoming), which also checked for Hawthorne effect, there are many small units in the treatment group. Because individual shops had little impact on the overall treatment effect, and there was barely any communication between shops, they had little incentive to exert effort beyond what their individual utility maximization required. Second, as we were informed by the firm's management, a number of pilot marketing initiatives (product campaigns, charity appeals, etc.) had been introduced before our team bonus scheme without being rolled out. With pilot schemes coming and going, there was little reason for the workers to expect this particular scheme to continue beyond the clearly communicated end in June 2014. In fact, the top management, convinced that the treatment effect was genuine, decided in early June 2014 to roll out the bonus scheme in all shops, treatment and control, which decision was communicated to the district managers in the end of June.

The decision to roll out the scheme enables us to compare the treatment and control group shops' response to the bonus in July to September 2014, which will provide statistical evidence for the rest of our arguments against Hawthorne effect. Namely, we argue that, had this effect existed, the control group shops would have increased their sales in July to September by less than 3%, the treatment effect in April to June 2014.

To support our argument, we estimate the post-treatment "treatment effect" and the implied effect of the team bonus on sales in the control group. The team bonus introduction procedure in the rollout stage was the same as in March 2014 (recall section 5), except all shops received the same letter. Unfortunately, a major reassignment of district managers coincided with the start of the rollout, resulting in a new district manager for two-thirds of the shops. Given the importance of the district manager in communications concerning the bonus scheme, we confine our post-treatment analysis to the subsample of 63 shops that did not see the change of district manager. This constant district manager (CDM) subsample does not differ significantly from the rest in terms of either descriptive statistics or the estimated treatment effect in April to June 2014.

As one would expect, the post-treatment "treatment effect" on the CDM subsample is zero. While this result does not in itself preclude Hawthorne effect, it does suggest that there was nothing unusual in the treatment effect we found in April to June 2014. Focusing on the control group shops in the CDM subsample, we estimate the effect of the bonus scheme on sales in these shops under the assumption of a constant trend in their sales. That is, we assume that in the absence of the bonus the year-on-year change in sales in July to September 2014 would be the same as in July to September 2013, and then contrast the log actual (10.11) and assumed (10.07) sales. The implied effect of the bonus on sales for these shops, 4%, is not consistent with Hawthorne effect, because it would have been lower than the 3% we previously found.

While neither the institutional context nor data provide any evidence consistent with Hawthorne effect on worker effort, a possibility to be so affected still remains for district manager effort. For instance, district managers could benefit from a positive treatment effect in their district as a way to signal their ability to the top management. One would then expect the district managers to spend more time with the treatment shops than with control shops. However, from the May 2014 employee survey we learn that there is no difference in the frequency of district manager visits between the treatment and control shops (4 to 5 visits per month on average in both groups).

6.3 Gaming

As we mentioned in section 3.2, the step-wise bonus may lead to "gaming", for example, through calibrating sales effort so as to just pass the bonus threshold. Anecdotally, we find a number of shops failing to reach their target by trivial amounts (for instance, one shop failed to reach the target by 16 Euros, and another one by 8 Euros) – an observation not consistent with gaming. In support of this observation, we learned from interviews with the district managers that, although the sales figures were communicated to all teams on a weekly basis, sales staff found it hard to estimate the likelihood of reaching the target because the demand was volatile. In line with this argument, we find that the treatment effect does not vary significantly with pre-treatment sales volatility.

FIGURE 5 ABOUT HERE.

Figure 5 offers a more systematic perspective on the symptoms of gaming by showing histograms of the log deviations of the actual sales from the target for the control and treatment groups separately. (For better visibility, only cases with the deviations within ±10% are included.) As an indication for possible gaming, we observe 7.5% of cases with excess sales of between 0 and 0.5 percent in the treatment group and 4.5% in the control group. However, this difference is not strong enough evidence for gaming for four reasons. First, even though the peak in the frequency right after 0 is distinct for the treatment group, the Kolmogorov-Smirnov test does not reject the null equality of excess sales distributions in the treatment and control group once the treatment effect is subtracted from excess sales (p-value = 0.363). Second, there are no similarly prominent peaks at other cutoff points (1%, 2%, 3%, 4% excess sales). Third, gaming would imply not only a peak above the target but also a trough just below, which we do not see at any of the cutoff points. Fourth, there are more cases in the treatment group than in control with excess sales above 4.5%, a level at which no extra bonus is paid and gaming is unlikely (29.2% vs. 23.6% in the treatment period). In fact, a naive difference-in-difference calculation produces a borderline significant treatment effect of 0.076 on the frequency of excess sales above 4.5%. Summing up, the evidence for gaming is weak, and even if there is gaming it would explain little of the treatment effect we have found.

7. Discussion

7.1 Mechanisms

The extra 3% of sales in the treatment shops compared to control may have been achieved by serving more customers (extensive margin) or by selling more per customer (intensive margin), or combination of both. In this section, we dissect our estimated treatment effect along these margins.

Starting with the intensive margin, we find (Table 5, row 1) a nearly zero treatment effect on the sales per customer-visit. This finding implies that up-selling, even if attempted, would contribute little to overall sales. This impression is confirmed by the results of a mystery shopping tour we made in 140 randomly selected shops in our sample in May 2014 (capacity constraints prevented us from touring every shop). Our research assistants were instructed to act like ordinary customers and to buy the “bread of the month” or the closest substitute to it. After leaving the shop, they were asked to take note of whether the question “Would you like anything else?” or similar was asked. We found that the frequency of asking the “anything else?” question was only slightly higher in the treatment group (79%) than in control (72%), a statistically insignificant difference. Furthermore, we found neither a significant correlation between asking this question and log sales in May, nor any part of the treatment effect disappearing once we include this question as control in our baseline regression.

TABLE 5 ABOUT HERE

Turning to the extensive margin, we observe in Table 5 (row 2) that the treatment effect on the number of customer-visits is commensurate with that on sales: 2.7% vs. 3%. Hence, the treatment effect occurs predominantly on the extensive margin. We analyze several channels through which this effect may have occurred. The first is extending opening hours by opening shop earlier or closing later. This cannot be done on Monday to Saturday in 95% of the shops because they are located on premises of large supermarkets and are forced by their rental agreements to exactly follow their host's opening hours. On Sunday, when supermarkets must be closed by law, bakeries may extend their hours. However, removing the 30% of shops in our sample that are open on Sunday and could therefore work longer then, does not change our results.

Another possibility to sell more is to over-order products from the central warehouse – at the cost of higher share of unsold goods. However, the automated ordering system gives little room for flexibility in orders. There is no treatment effect on the share of unsold goods, either (Table 2).

Extra customer visits could also have been achieved by offering better, friendlier customer service. To test this possibility, we asked our research assistants on the May 2014 mystery tour to evaluate shop staff friendliness on a Likert scale. Surprisingly, their evaluations, either with or without mystery shopper fixed effects, are *negatively* correlated with sales, which goes against the hypothesis that friendlier customer service is behind the observed treatment effect.

The only channel we are left with is improved operational efficiency within the existing operational constraints (opening hours, product ordering rules, standards of service, etc). This could be achieved through a combination of i) reallocating work shifts to better match labor input with demand, ii) shop manager ability to manage their employees under given shifts, or iii) working faster under given shift allocation and management input. Our data do not support the first mechanism - work shift allocation - because the treatment effect in April is the same as in May or June (Table 4); yet, the work shifts for April were planned one month before the treatment was communicated.

Turning to the role of management input, we know already that it is not the only channel behind the effect of team bonus on sales, because the treatment effect interacts substantially with worker characteristics, most notably, the share of un-incentivized mini-jobbers (Table 4). To test the contribution of management input, we allow the treatment effect to depend on its several proxies, all measured before the bonus scheme was introduced in April 2014. The proxies are: shop manager monthly working hours, tenure, average bonus he or she received between January 2012 and March 2014, and the linear combination of the above three proxies with weights estimated from the production function regression of shop sales on shop, worker and manager characteristics. None of our shop manager input measures differ between the treatment and control groups, and none interacts significantly with the treatment effect. Thus, while its role in generating sales cannot be denied, there are no signs that shop manager input significantly shapes the magnitude of the effect of team incentives on sales.

We are left with working faster as the only remaining mechanism facilitating the operational efficiency gains behind the treatment effect. Shop assistants' being quicker at routine tasks, such as cleaning or delivery, so that they could spend more time at the counter, is an illustration of "working faster".

7.2 The 'political economy' of new management practice implementation in firms

Our statistical findings as well as first-hand experience in implementing the team bonus scheme in the firm provide an insider perspective on the adoption of management practices by firms, an issue much discussed in the organizational economics literature. The big question is: why do some firms adopt productivity-enhancing management practices while other, even though in the same industry, do not? The literature came up with several answers, among which most frequently discussed are lack of knowledge (Bloom et al., 2013), heterogeneity in management practices' performance effects and limited organizational capabilities (Bandiera et al., 2011; Ichniowski and Shaw, 2012), and product market competition (Bloom and Van Reenen, 2010; Syverson, 2011; Bloom et al., 2014), which is arguably most important as it drives firms to try new practices despite the above.

Our findings speak to all these answers. It was lack of awareness ("[Monetary incentives to sales staff] were simply never on our agenda.") that prevented the firm from adopting sales staff incentives earlier. Significant heterogeneities in the team bonus' effect that we detected even within the same firm would not make the provision of incentives worthwhile to some of the firms. Our experience of communicating with the firm revealed several limitations on the resources the firm's employees were able to commit to new projects given their other responsibilities. However, it was the product market competition, intensified by the entry of discounter supermarkets, that pushed our firm to think harder about its HR management practices and implement our proposal despite the extra effort it required of them.

An additional contribution our study makes to the literature on management practices adoption is highlighting the importance of internal politics - even in the presence of intense competition which should overcome any partisanship. There may be tensions between the new and existing management practices, causing resistance to change. It is important to provide mechanisms to relieve these tensions. Two instances of the conflict between new and existing practices apply in our case. First, team bonus for sales staff would certainly imply higher personnel costs, whereas its sales benefits were not clear at the beginning; hence the resistance of management, whose bonuses depended on sales and personnel costs, to the team bonus (section 3.1). It took a commitment of the CEO to allocate a separate budget for the team bonus to overcome this resistance. Second, while some employees stood to gain from team bonus, others would loose. The case in point are HR personnel who would have to do more work administering the bonus without directly benefiting from it. We took over much of the administrative effort (e.g., printing information leaflets, training district managers), thus easing the resistance of the HR workers to our new practice.

7.3 Practical aspects of management practice implementation

Here we share a few reflections on List's (2011) authoritative guide for field experimenters, which inspired many elements of our own experimental design. In particular, we employed economic theory to inform our treatment and interpret the findings; we spent effort randomizing and measuring the statistical power of our experiment according to the current best scientific practice; managed to find a champion for our cause in the top management; and addressed organizational complications to our experiment. In the process, we gained experience in finding ways to address List's guidelines, which may be useful for researchers and practitioners interested in implementing, and measuring the effect of, new management practices in firms.

Given the existing field experimental guidelines, our most instructive experience is in dealing with organizational resistance to new practice implementation (tip 6 in List, 2011). To remind, the organizational resistance to our ``pilot'' was due to conflicting incentive schemes for managers (existing) and sales staff (new) and due to extra burden on the HR personnel. To address these two causes for resistance, we managed to allocate a separate budget for the team bonus, and took over some of the administrative work. Field experimenters would therefore do well by anticipating possible negative externalities from implementing new practices, and by organizing resources necessary to minimize these externalities.

Trust between the experimenter and the firm is essential for gaining resources to run a successful experiment. To gain trust, List (2011, p.13) recommends building up a record of research engagement with the firm prior to the experiment of main interest to the researcher. In addition to having an early success with our study company on another project some time before the team incentives, we built trust through constant communication with managers at all levels of hierarchy, and through recruiting the workers council on our side. The workers council support was crucial for allowing us to go ahead with our experiment, as well as for our unequal treatment to gain legitimacy with the control group workers should they come to know about it. Our experience with the workers council is instructive: it suggests that institutions that are unlikely to support experimentation, when convinced, will help the experimenter by boosting trust and legitimacy.

8. Conclusion

Teams are a ubiquitous feature of modern production, and so are monetary incentives. While the knowledge about the effectiveness of individual incentives is both broad and deep, much less is known about team incentives. Problems of endogeneity, complementarities and self-selection into teams make causally interpretable evidence about the effectiveness of team incentives hard to obtain. We contribute to the incentives literature by providing causal evidence on the effectiveness of team incentives. We have conducted a large-scale natural field experiment involving 193 shops and 1,300 employees of a bakery chain in Germany. Our estimated treatment effect is around 3%, or one third of the sales' standard deviation. There is also substantial heterogeneity, with the treatment effect being largest in big towns, shops with younger workforce and few mini-jobbers. The latter finding suggests effort complementarities within teams. The single most important immediate cause of the treatment effect is increased customer traffic; there is no effect on sales per customer visit. Improved operational efficiency gain through working faster is the most plausible mechanism behind the treatment effect.

References

Alchian, Armen A., and Harold Demsetz. "Production, information costs, and economic organization," *American Economic Review,* 62 (5) (1972), 777-795.

Allen, Natalie. J., and John P. Meyer, "The measurement and antecedents of affective, continuance and normative commitment to the organization," *Journal of Occupational and Organizational Psychology*, 63 (1) (1990), 1-18.

Auriol, Emmanuelle, Guido Friebel, and Lambros Pechlivanos, "Career concerns in teams," *Journal of Labor Economics*, 20 (2) (2002), 289-307.

Baily, Martin Neil, and Robert M. Solow, "International productivity comparisons built from the firm level," *Journal of Economic Perspectives*, 15 (3) (2001), 151-172.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Field Experiments with Firms," *Journal of Economic Perspectives*, 25 (3) (2011), 63-82.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Team Incentives: Evidence from a Field Experiment," *Journal of the European Economic Association*, 11 (5) (2013), 1079-1114.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Connections and Incentives in the Workplace: Evidence from Personnel Data," *Econometrica*, 77 (4) (2009), 1047-1094.

Barankay, Iwan (2014), "Rank Incentives: Evidence from a Randomized Workplace Experiment," *Working Paper*.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How much should we trust difference-in-difference estimates?" *Quarterly Journal of Economics*, 119 (1) (2004), 249-275

Bertrand, Marianne, and Francis Kramarz, "Does entry regulation hinder job creation? Evidence from the French retail industry," *Quarterly Journal of Economics,* 117 (4) (2002), 1369-1413.

Barrios, Thomas, "Optimal stratification in randomized experiments," Mimeo, Harvard University (2014).

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, Christos Genakos, Raffaella Sadun, and John Van Reenen, "Management Practices across Firms and Countries," *Academy of Management Perspectives*, 26 (1) (2012), 12-33.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen, "The new empirical economics of management," *Journal of the European Economic Association*, 12 (4) (2014), 835-876.

Bloom, Nicholas, James Liang, John Roberts, and Zhichung Jenny Ying, “Does working from home work? Evidence from a Chinese experiment,” *Quarterly Journal of Economics,* forthcoming.

Bloom, Nicholas, and John Van Reenen, “Measuring and explaining management practices across firms and countries,” *Quarterly Journal of Economics*, 122 (4) (2007), 1351-1408.

Bloom, Nicholas, and John Van Reenen, “Why Do Management Practices Differ across Firms and Countries?” *Journal of Economic Perspectives*, 24 (1) (2010), 203-224.

Bloom, Nicholas, and John Van Reenen, “Human resource management and productivity,” *Handbook of Labor Economics,* 4B (2011), 1697-1767.

Boning, Brent, Casey Ichniowski, and Kathryn Shaw, “Opportunity counts: Teams and the effectiveness of production incentives,” *Journal of Labor Economics*, 25 (4) (2007), 613-650.

Burgess, Simon, Carol Propper, Marisa Ratto, Stephanie von Hinke Kessler Scholder, and Emma Tominey, “Smarter Task Assignment or Greater Effort: the Impact of Incentives on Team Performance,” *Economic Journal* 120 (2010), 968-989.

Cameron, A. Colin, and Douglas L. Miller, “A practitioner's guide to cluster-robust inference,” Forthcoming in *Journal of Human Resources* (2015).

Courty, Pascal, and Gerald Marschk, “Measuring government performance: Lessons from a federal job-training program,” *American Economic Review*, 87 (2) (1997), 383-388.

Delfgaauw, Josse, Robert Dur, Arjan Non, and Willem Verbeke, “Dynamic incentive effects of relative performance pay: A field experiment,” *Labour Economics*, 28 (1) (2014), 1-13.

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.

Englmaier, Florian, Andreas Roider, and Uwe Sunde (2014), “The Role of Communication of Performance Schemes: Evidence from a Field Experiment,” *Working Paper*.

Esteban, Joan, and Debraj Ray, “Collective action and the group size paradox,” *American Political Science Review*, 95 (3) (2001), 663-672.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan, “Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation,” *Journal of Political Economy*, 111 (3) (2003), 465-497.

Harrison, Glenn W., and John A. List, “Field Experiments,” *Journal of Economic Literature,* 42 (4) (2004), 1009-1055.

Holmström, Bengt, “Moral hazard in teams,” *The Bell Journal of Economics*, 13 (2) (1982), 324-340.

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

Ichniowski, Casey, and Kathryn L. Shaw, “Insider Econometrics: A Roadmap to Estimating Models of Organizational Performance,” *Handbook of Organizational Economics*, (2012), Robert Gibbons and John Roberts, eds., Princeton University Press.

Iranzo, Susana, Fabiano Schivardi, and Elisa Tosetti, “Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data,” *Journal of Labor Economics*, 26 (2) (2008), 247-285.

Itoh, Hideshi, “Incentives to help in multi-agent situations,” *Econometrica*, 59 (3) (1991), 611-636.

Kandel, Eugene, and Edward P. Lazear, “Peer Pressure and Partnerships,” *Journal of Political Economy*, 100 (4) (1992), 801-817.

Lazear, Edward. P., “Performance Pay and Productivity,” *American Economic Review*, 90 (5) (2000), 1346-1361.

List, John (2011), “Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off,” *Journal of Economic Perspectives*, 25 (3) (2011), 3-16.

List, John, and Imran Rasul, “Field Experiments in Labor Economics,” *Handbook of Labor Economics*, 4A (2011), 103-228.

Mas, Alexandre, and Enrico Moretti, “Peers at Work,” *American Economic Review*, 99 (1) (2009), 112-45.

Nalbantian, Haig R., and Andrew Schotter, “Productivity under Group Incentives: An Experimental Study,” *American Economic Review*, 87 (3) (1997), 314-341.

Shearer, Bruce, “Piece rates, fixed wages and incentives: Evidence from a field experiment,” *Review of Economic Studies*, 71 (2) (2004), 513-534.

Stigler, Stephen M., “Regression towards the mean, historically considered,” *Statistical Methods in Medical Research*, 6 (2) (1997), 103-114.

Syverson, Chad, “What determines productivity?” *Journal of Economic Literature*, 49 (2) (2011), 326-365.

Tables and Figures

Figure 1: The Team Bonus

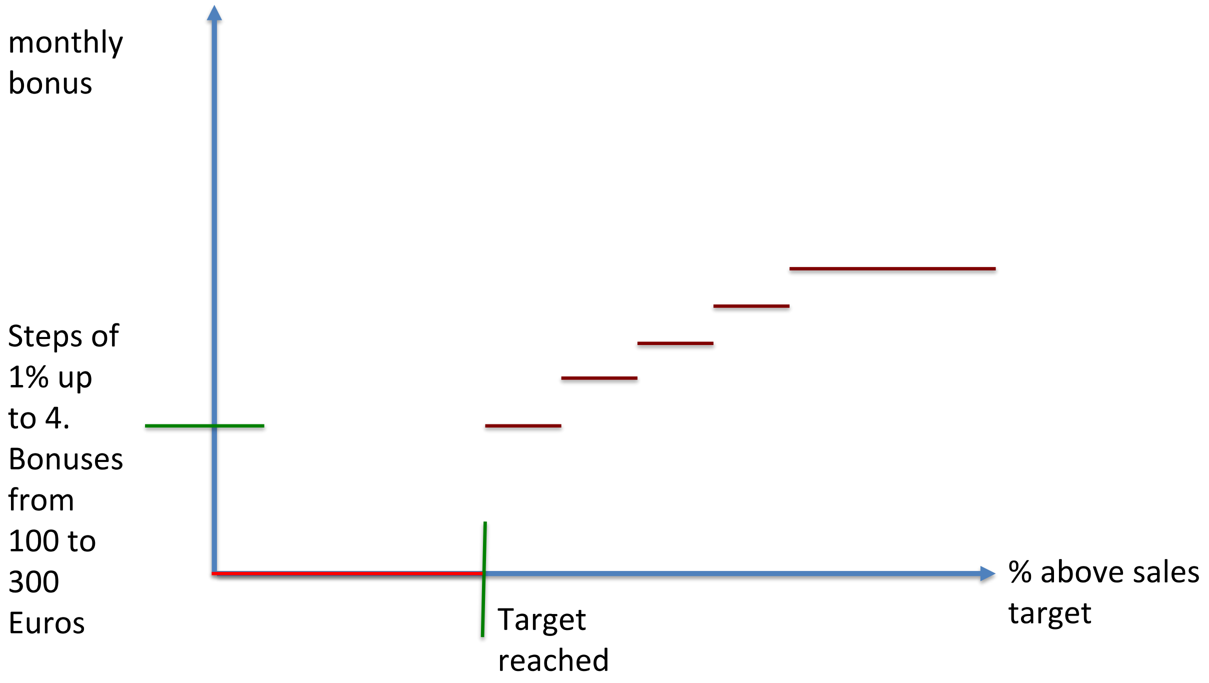


Figure 2: A map of shops by treatment (white) and control group (black)

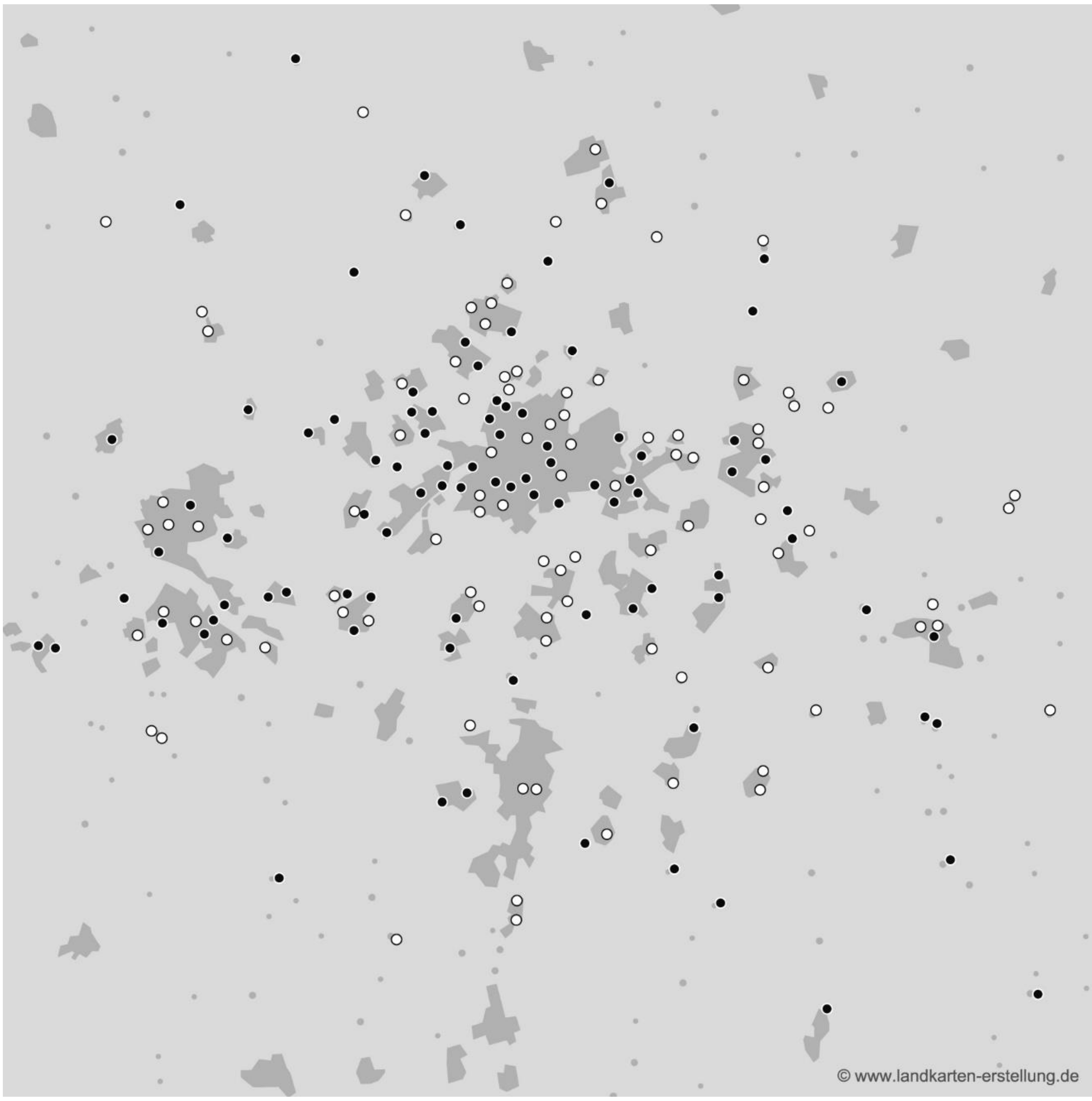


Figure 3: Scatter plot, year on year sales growth on log sales, April, May and June 2013

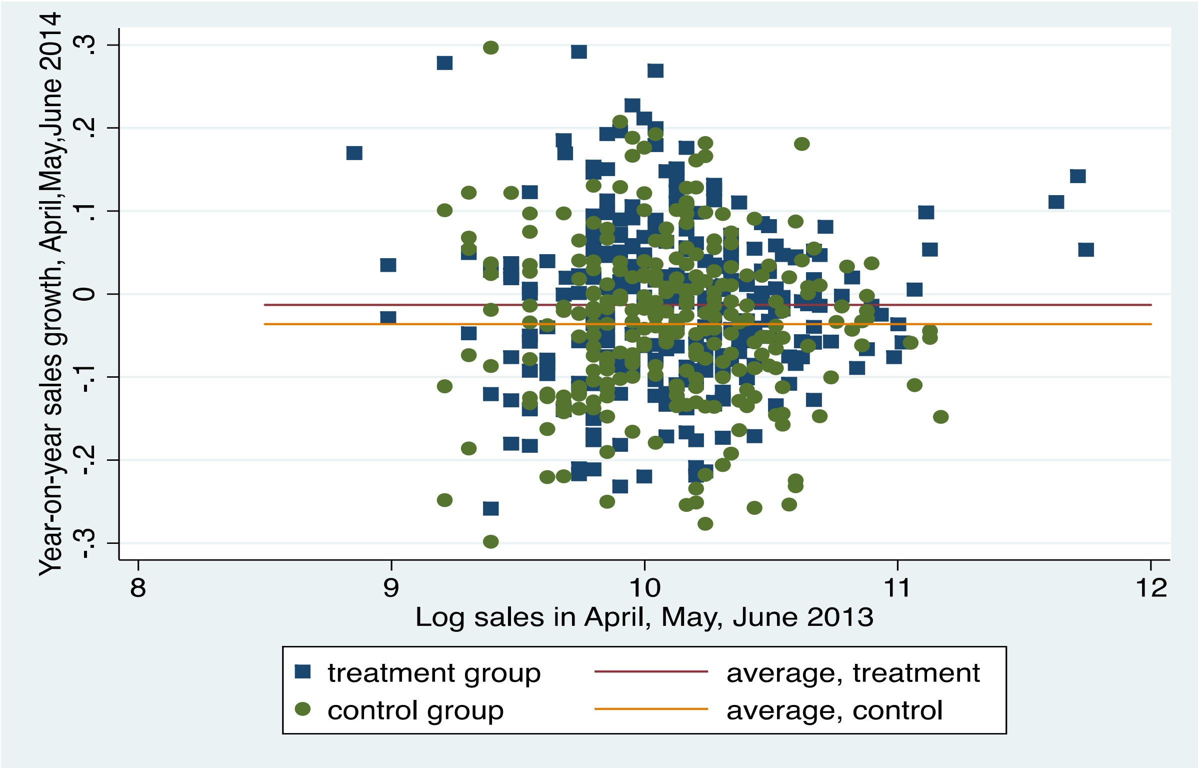


Figure 4: Kernel distribution sales growth treatment versus control group

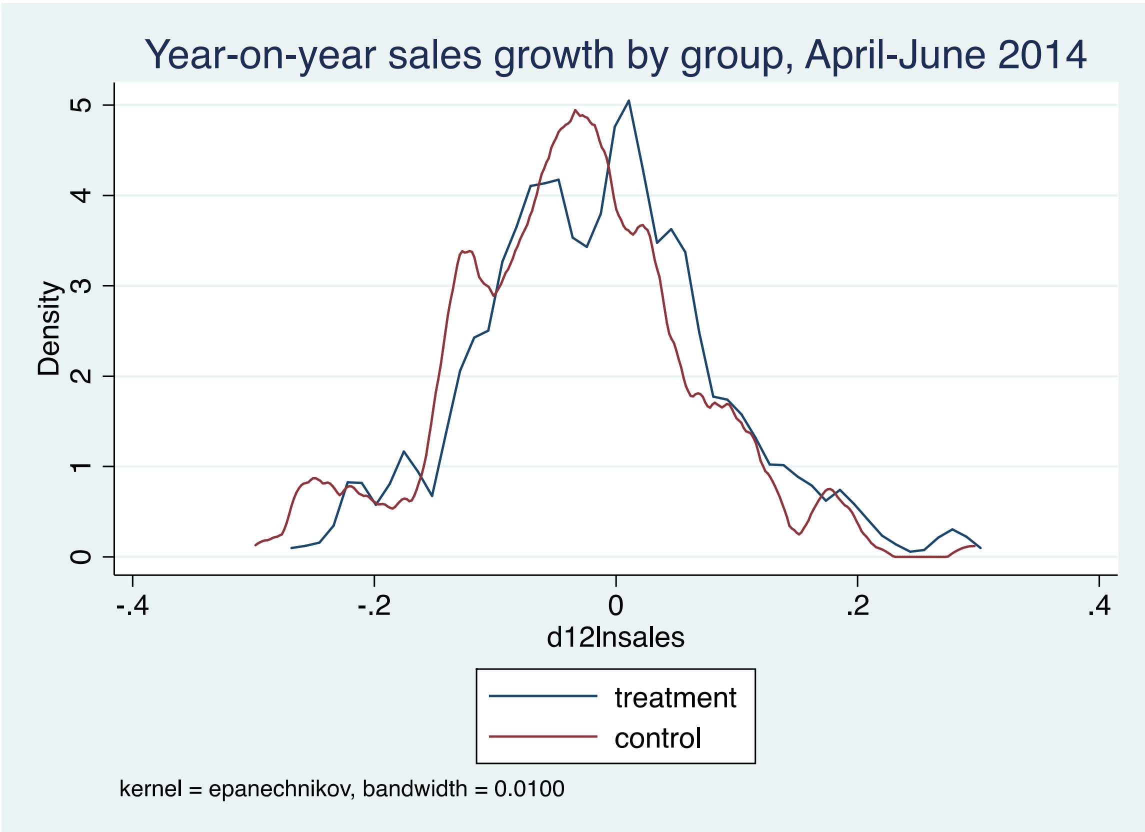
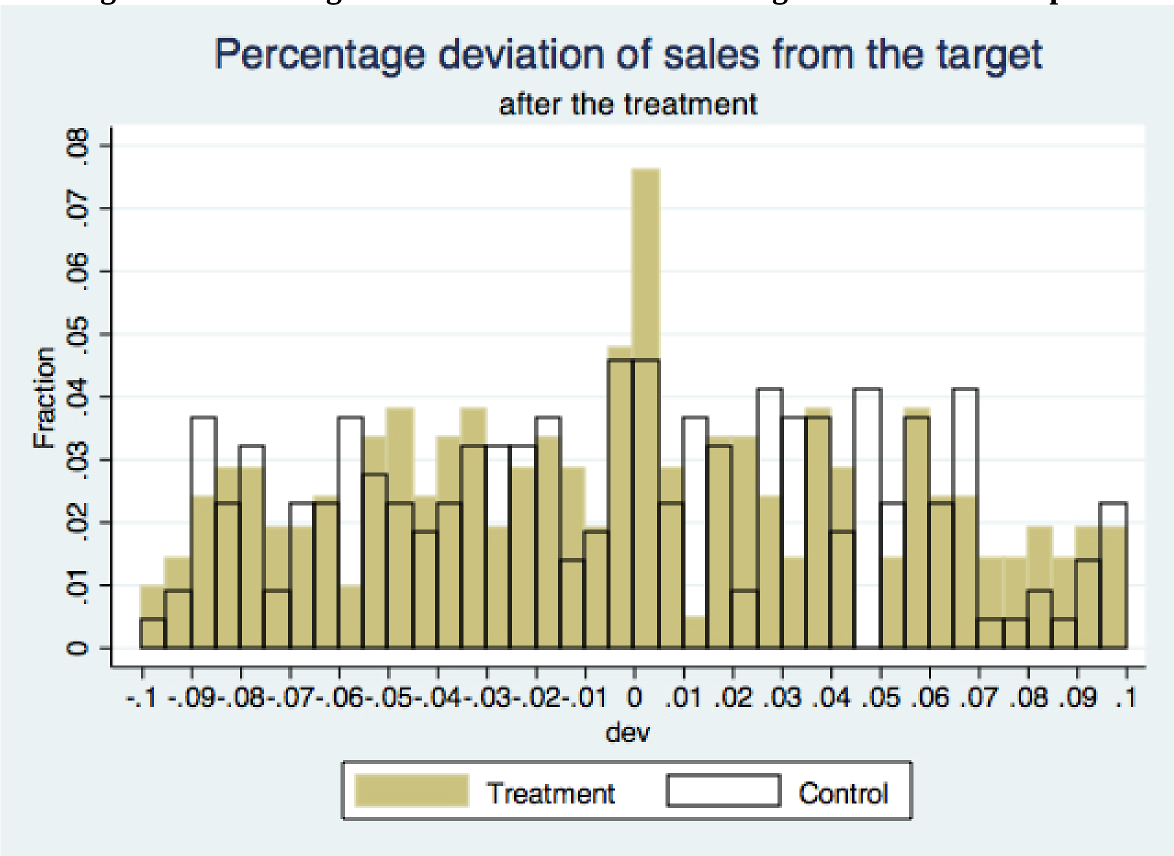


Figure 5: Percentage deviation of sales from the target in the treatment period



Note: for better visibility only deviations within ±10% are included.

Table 1: Characteristics of the control and treatment shops before the treatment

	Control (n = 96)	Treatment (n = 97)	t-test p-value
	Panel A: Quantitative performance indicators		
Mean monthly sales (SD)	27453 (11481)	28194 (14542)	0.695
Mean monthly sales (in logs, SD)	10.14 (0.39)	10.15 (0.41)	0.846
Unsold goods as % of sales (SD)	16.16 (7.0)	15.54 (6.9)	0.331
Mean number of customer-visits (SD)	10028 (3921)	10131 (4018)	0.856
Mean monthly quit rate (SD)	1.9% (4.1%)	1.8% (4.1%)	
Frequency of achieving the sales target	35.8%	35.2%	0.860
	Panel B: Qualitative performance indicators		
Mean mystery shopping score 2013 (SD)	96.1%	95.5%	
Mean mystery shopping score 2014 (SD)	97.6%	97.6%	
	Panel C: Shop location		
Big town	37.6%	33.6%	
Medium/small town	26.0%	29.6%	
Village	36.4%	36.7%	
	Panel D: Characteristics of shop managers		
Mean age, years	39.8 (6.4)	40.9 (6.3)	
Share of females	94.9%	93.0%	
Share of full-time employees	71.8%	64.8%	
	Panel E: Characteristics of sales agents		
Total number of sales agents	552	580	
Mean number of agents per shop (SD)	7.4 (3.2)	7.4 (3.2)	
Mean age, years	39.5 (6.1)	39.9 (6.0)	
Share of females	93.1	92.4	
Share of employees with a permanent contra	66.6%	67.9%	
Share of full-time employees	9.7%	10.4%	
Share of part-time employees	56.7%	59.7%	
Share of mini-jobbers	33.6%	29.9%	
Share of unskilled workers	77.5%	72.3%	
	Panel F: Employee attitudes		
Mean commitment score (SD)	4.69 (1.38)	4.68 (1.42)	0.895
Mean work satisfaction score (SD)	4.61 (1.37)	4.55 (1.31)	0.547
Mean overall satisfaction score (SD)	5.15 (1.46)	5.15 (1.38)	0.997

Standard deviations are in parentheses. Column 3 reports the p-values of the two-sided t-test of equality of the means for a selection of variables. "Big town", "medium/small town" and "village" refer to municipalities with more than 90,000; 5,000 to 60,000; and fewer than 5,000 inhabitants, respectively. Panels D and E are based on the personnel records from the firm as of July 1 2014, excluding apprentices and interns (18 in the control and 11 in the treatment group). Panel F reports the means of the commitment, work satisfaction and overall satisfaction scores constructed according to Allen and Meyer (1990) from the employee survey administered in March 2014. In total, 563 employees in the control, and 580 employees in the treatment group participated in the survey (response rate 79.5%).

Table 2: Characteristics of the control and treatment shops in the treatment period (April - June 2014)

	Panel A: Quantitative performance indicators		
	Control (n = 96)	Treatment (n = 97)	t-test p-value
Mean monthly sales (SD)	25376 (10708)	26995 (15036)	0.061
Mean monthly sales (in logs, SD)	10.06 (0.40)	10.10 (0.42)	0.034
Unsold goods as % of sales (SD)	22.88 (9.8)	22.35 (13.3)	0.940
Mean number of customer-visits (SD)	9115 (3582)	9465 (3790)	0.062
Mean monthly quit rate (SD)	1.42% (4.89)	1.69% (5.64)	0.336
Frequency of achieving the sales target	44.8%	49.1%	0.442
	Panel B: Qualitative performance indicators		
Mean mystery shopping score (SD)	98.2%	97.6%	0.295
	Panel C: Employee attitudes		
Mean commitment score (SD)	4.56 (1.28)	4.62 (1.33)	0.570
Mean work satisfaction score (SD)	4.39 (1.34)	4.48 (1.20)	0.418
Mean overall satisfaction score (SD)	4.86 (1.36)	4.99 (1.33)	0.233

Column 3 reports the p-values of the two-sided significance test for the difference-in-difference estimate of the treatment effect. The second employee survey was administered in May 2014 with a response rate of 76%.

Table 3: Treatment effect estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment effect	0.032	0.033	0.030	0.030	0.032	0.026	0.027
	(.013)	(.011)	(.014)	(.014)	(.014)	(.014)	(.014)
Shop fixed effects	Yes	Yes	Yes	No	No	No	No
Month dummy variables	Yes	Yes	No	No	No	No	No
Other controls	Yes	Yes	No	Yes	Yes	Yes	Yes
Observations	49316	4904	386	193	577	561	577

The table shows the difference-in-difference treatment effect estimates based on several regression specifications with the log sales as the dependent variable. In all specifications the unit of observation is individual shop. In specification 1, we regress monthly sales from January 2012 until June 2014 on the "treatment group" and "after treatment" dummies and their cross-product. Specification 2 is the same but omits the outliers, defined as year-on-year sales change exceeding 30% (roughly the top and bottom 1% of the sales growth distribution). The reasons for such substantial increases or decreases in sales are construction sites close to the bakeries, competitors who enter or leave the market, temporary closures of shops because of renovations or sunny weather, which affects sales in bakeries located in shopping centers. Specification 3 is the same as 1, except that we use log average sales over the periods before and after the treatment (hence two observations per shop). Specification 4 includes past sales as an additional control, hence one observation per shop. In specification 5, we regress the log monthly sales in April, May and June 2014 (the treatment period) on the treatment dummy and the baseline sales in the respective shop, defined as the log average sales over the pre-treatment period. In specification 6, we regress the log monthly sales in the treatment period on the treatment dummy and the log sales in the respective months in 2013. Specification 7 is the same as 5 except that we use the log average sales in January-Mach 2014 as the baseline. Standard errors are clustered by shop. Cluster-bootstrapped standard errors (available on request) are similar in magnitude.

Table 4: Treatment effect heterogeneity

	Whole sample	Big towns	Midsize towns	Villages
		Panel A: Treatment effect by shop location		
	Whole sample	Big towns	Midsize towns	Villages
April 2014	0.029	0.059	0.023	0.004
	(.011)	(.019)	(.018)	(.019)
May 2014	0.037	0.055	0.049	0.011
	(.022)	(.051)	(.023)	(.020)
June 2014	0.029	0.049	0.045	-0.001
	(.014)	(.024)	(.026)	(.022)
Overall	0.032	0.055	0.038	0.005
	(.013)	(.025)	(.020)	(.019)
	Panel B: Treatment effect by quartile of shop size (number of workers)			
Quartile 1	0.001	0.016		
	(.024)	(.041)		
Quartile 2	0.022	0.005		
	(.022)	(.038)		
Quartile 3	0.041	0.046		
	(.027)	(.054)		
Quartile 4	0.059	0.125		
	(.025)	(.043)		
	Panel C: Treatment effect by shop-average employee age			
Above median	0.001	0.043		
	(.017)	(.031)		
Below median	0.061	0.063		
	(.019)	(.036)		
	Panel D: Treatment effect by the average share of mini-job employees			
Quartile 1 (<0.06)	0.071	0.052		
	(.033)	(.076)		
Quartile 2 (0.06-0.11)	0.050	0.098		
	(.026)	(.042)		
Quartile 3 (0.11-0.16)	0.003	0.053		
	(.019)	(.039)		
Quartile 4 (>0.16)	-0.003	-0.019		
	(.021)	(.035)		
	Panel E: Treatment effect by pre-treatment deviation of sales targets			
	E1: Distance measure: pre-treatment average sales/target difference			
Quartile 1 (<-8%)	0.046			
	(.026)			
Quartile 2 (-8% to -4.5%)	0.036			
	(.028)			
Quartile 3 (-4.5% to 0%)	0.047			
	(.027)			
Quartile 4 (>0%)	0.003			
	(.017)			
	E2: Distance measure: pre-treatment frequency of achieving the target			
Quartile 1 (<16%)	0.052			
	(.022)			
Quartile 2 (16% to 30%)	0.048			
	(.025)			
Quartile 3 (30% to 50%)	0.026			
	(.030)			
Quartile 4 (>50%)	-0.009			
	(.016)			

The cells in the table give estimated treatment effect in a given month and location. The regression specification is the same as spec. 1 in Table 3. Standard errors are clustered by shop. For example, 0.059 is the treatment effect in April 2014 in shops located in big towns. Standard errors are clustered by shop. In Panel B, shop size is defined as the number of workers employed in a shop excluding those on a mini job. In Panel C, the samples are split into below and above the median age/tenure of the workforce excluding workers employed in a mini job. In Panel D, the share of mini-job workers is defined as the ratio of the hours worked by these workers to the total hours worked. Quartiles of the share of mini-job workers are very similar for every location, and so are defined on the whole sample.

Table 5: Treatment effect on the number of customer visits and sales per customer visit

	Whole sample	Big towns	Midsize towns	Villages
Treatment effect on sales per visit	0.004	0.008	0.004	0.000
	(.007)	(.018)	(.005)	(.007)
Treatment effect on customer visits	0.027	0.046	0.032	0.006
	(.011)	(.020)	(.019)	(.017)

Appendix I

Information leaflet

<LOGO OF THE BAKERY>

AN ALLE VOLL- UND TEILZEITKRÄFTE: VERDIENEN SIE SICH IHREN TEAM-BONUS

In den Monaten April, Mai und Juni 2014 erhält das Team Ihrer Filiale einen Team-Bonus bei Erreichung oder Übererfüllung der Umsatzziele.
So sieht das Bonus-Programm für Voll- und Teilzeitkräfte aus:

- Bei Erreichung oder Übererfüllung **von bis zu 1%**, erhält das Filial-Team einen Bonus von **100€** für den entsprechenden Monat.
- Bei **1% bis 2%** über dem Umsatzziel erhält das Filial-Team einen Bonus von **150€**.
- Bei **2% bis 3%** beträgt der Team-Bonus **200€**.
- Bei **3% bis 4%** beträgt der Team-Bonus **250€**.
- Bei **4% oder mehr** gibt es einen Team-Bonus von **300€**.

Jedes Filial-Team kann also im Quartal einen **Bonus von bis zu 900€** erreichen!

Bitte beachten Sie:

- Details zur Aufteilung unter den Team-Mitgliedern und Fehlzeiten finden Sie im Infobrief.
- Leider können wir diese Regelung aus steuerrechtlichen Gründen nicht für geringfügig Beschäftigte anwenden.

Bei Fragen wenden Sie sich bitte an Ihre Bezirksleiter/innen, die Ihnen gerne weiterhelfen und ihnen regelmäßig mitteilen werden, ob sie Ihre Umsatzziele erreicht haben.

Appendix II: Proofs of the model's predictions

Prediction 3: Individual effort decreases with team size N if effort complementarities are not too strong ($p \gg 1/2$). However, depending on the strength of effort complementarities and the convexity of the costs of effort function, the team's *total* effort may increase or decrease with N.

Assuming, as before $\left|\Phi''\left(aN^{\frac{1}{p}}\right)e^* - y_0\right| \ll \Phi'(aN^{\frac{1}{p}}e^* - y_0)$,

$$\frac{de^*}{dN} = -\frac{a \cdot BN^{\frac{1-3p}{p}}}{p} * \frac{(1-2p)\Phi'\left(aN^{\frac{1}{p}}e^* - y_0\right) + N^{\frac{1}{p}}e \cdot a\Phi''\left(aN^{\frac{1}{p}}e^* - y_0\right)}{\frac{d^2\pi}{de_i^2}} < 0$$

when $p \gg 1/2$. For the total effort,

$$\frac{d(Ne^*)}{dN} = e^* + N\frac{de^*}{dN}$$

$$= e^* - \frac{a \cdot BN^{\frac{1-2p}{p}}}{p} * \frac{(1-2p)\Phi'\left(aN^{\frac{1}{p}}e^* - y_0\right) + N^{\frac{1}{p}}e \cdot a\Phi''\left(aN^{\frac{1}{p}}e^* - y_0\right)}{\frac{d^2\pi}{de_i^2}},$$

whose sign is ambiguous. It can be shown that when output is linear in effort (no complementarities, p = 1), $\Phi(x) \approx x$, and the costs of effort are quadratic, the negative effect of N on individual effort is exactly offset by gains in the total effort, giving $\frac{d(Ne^*)}{dN} = e^*$=0 (see also Esteban and Ray (2001) for the same result). Normalising quantities to suppress the inessential parameters a, b, B and y_0,

$$\pi(e_i, e_{-i}) = \frac{1}{N}\left(e_i + \sum_{j \neq i} e_j\right) - e_i^2$$

Maximizing π assuming an interior solution, we obtain $e^* = \frac{1}{2N}$ and $\sum e^* = \frac{1}{2}$, which does not depend on N. More generally, approximating $\Phi(x) = x^\gamma$ and $c(x) = x^k$,

$$\pi(e_i, e_{-i}) = \frac{1}{N}\left(e_i^p + \sum_{j \neq i} e_j^p\right)^{\frac{\gamma}{p}} - e_i^k, k > 1$$

$$N \cdot e^* = \left(\frac{\gamma}{k}\right)^{\frac{1}{k-\gamma}} \cdot N^{\frac{\gamma-2p}{p(k-\gamma)}+1} \tag{7}$$

The sign of the exponent of N in (7) determines the relationship between total effort and team size: it is positive when $k > \gamma + 2 - \gamma/p$, and negative otherwise.

Prediction 5: Team effort decreases with the share of non-incentivized members in the team.

Letting θ be their share, the individual payoff function is

$$\pi(e_i, e_j) = \frac{1}{N} B\Phi\left(a\left(e_i^p - \theta(e_i^p - e_0^p) + \sum_{j\neq i}\left(e_j^p - \theta(e_j^p - e_0^p)\right)\right) - y_0\right) \qquad (8)$$
$$-b \cdot c(e_i),$$

The first-order condition for an interior solution for the incentivized workers' effort $e^* > e_0$ is

$$\left.\frac{d\pi}{de_i}\right|_{e_i=e^*} = aN^{\frac{1-2p}{p}} B\Phi'\left(aN^{\frac{1}{p}}\left(e^{*p} - \theta(e^{*p} - e_0^p)\right)^{\frac{1}{p}} - y_0\right) \qquad (9)$$
$$\cdot\left(e^{*p} - \theta(e^{*p} - e_0^p)\right)^{\frac{1-p}{p}} e^{*p-1}(1-\theta) - b \cdot c'(e^*) = 0$$

One can see immediately that the expression in (9) falls in θ.

Prediction 6: The effort under the bonus will depend on the frequency of reaching the targets in the past, without the bonus. More successful teams' effort response to the bonus will be weaker than that of less successful teams. However, depending on the costs of effort, extremely unsuccessful teams may not respond to the bonus at all, choosing the corner solution $e^* = e_0$ instead.

To see this, assume that without the bonus every member of the team puts in the minimum acceptable effort e_0. Then the success in reaching the target is determined by y_0. Consider first the interior solution case, when $e_0 < e^* < e_{max}$.

$$\left.\frac{de^*}{dy_0}\right|_{e^*=e_0} = \frac{aN^{\frac{1-2p}{p}} B\Phi''\left(aN^{\frac{1}{p}} e_0 - y_0\right)}{\frac{d^2\pi}{de_i^2}},$$

which is positive when $aN^{\frac{1}{p}} e_0 < y_0$ (except output at e_0 falls behind the target), and negative otherwise. Thus, the more successful a team has been, the less effort it will put in a given bonus. However, the corner solution $e^* = e_0$ may be chosen by very unsuccessful team when, although $\left.\frac{de^*}{dy_0}\right|_{e^*=e_0} > 0$ given their record, the positive marginal benefit of effort is too small to offset the marginal costs (recall the first-order condition (6)). Whether the corner solution will occur depends on the costs of effort.