I Z A

IZA DP No. 9316

**Team Incentives and Performance:
Evidence from a Retail Chain**

Guido Friebel
Matthias Heinz
Miriam Krüger
Nick Zubanov

August 2015

Team Incentives and Performance: Evidence from a Retail Chain

Guido Friebel
Goethe University Frankfurt, CEPR and IZA

Matthias Heinz
University of Cologne and CEPR

Miriam Krüger
Goethe University Frankfurt

Nick Zubanov
Goethe University Frankfurt and IZA

ABSTRACT

Team Incentives and Performance: Evidence from a Retail Chain[1]

We test the effectiveness of team incentives by running a natural field experiment in a retail chain of 193 shops and 1,300 employees. As a response to intensified product market competition, the firm offered a bonus to shop teams for surpassing sales targets. A bonus to teams rather than individuals was a natural choice because the firm does not measure individual performance and relies on flexible task allocation among employees. On average, the 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 for each bonus dollar an extra $3.80 of sales, and $2.10 of operational profit. The results show the importance of complementarities within teams and suggest that improved operational efficiency is the main mechanism behind the treatment effect. Our analysis of heterogeneous treatment effects offers a number of insights about the anatomy of teamwork. The firm decided to roll out the bonus to all of its shops, and the performance of treatment and control shops converged after the roll-out.

JEL Classification: J3, L2, M5

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

Corresponding author:

Guido Friebel
Goethe University Frankfurt
Chair of Human Resource Management
P.O. Box 52
Grüneburgplatz 1
60629 Frankfurt
Germany
E-mail: gfriebel@wiwi.uni-frankfurt.de

1. Introduction

"How can members of a team be rewarded and induced to work efficiently?" This classical 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 for the principal, team incentives should be expected to be weakened by free-riding (Holmström, 1982).

From a broader theoretical point of view, team incentives have unclear effects on output because much depends on technological assumptions. If efforts simply add up to an outcome plus noise, a team incentive translates into an individual incentive that is proportional to team size. This (weak) individual incentive, then, determines individual effort choices. However, in the presence of complementarities, strategic behavior of team members is more complex, and there is scope for multiple, high- and low-effort equilibria (Cooper and John, 1988). Hence, even without considering the important point of peer pressure between members of a team as in Kandel and Lazear (1992), or the many potential behavioral effects that may be going on within teams (Burks et al., 2009; Mohnen et al., 2008; Friebel and Schnedler, 2010; Kosfeld and von Siemens, 2014), the question of whether or not team incentives increase performance is mainly an empirical one. However, in stark contrast to *individual* incentives that have been shown to work well in the field (Lazear, 2000; Shearer, 2004; Bandiera et al., 2009), the jury on the effectiveness of *team* incentives is still out (Bloom and Van Reenen, 2011).

We bridge the gap in the evidence on the effectiveness of team incentives by running a randomized experiment in which a team bonus was introduced in half of the 193 bakery shops of a German retail chain during the period between April and June 2014. A team rather than an individual bonus was used, because the firm only measures the performance of shops, not of individual sales assistants. The shops have always operated in a teamwork organization in which workers (on average seven per shop) carry out a variety of tasks, such as handling the goods delivered, operating the oven, serving customers, and operating the cash register. The volatility of demand makes specialization expensive (Friebel and Yilmaz, 2015), as agents would be idle much of their time. Employees are expected to help each other; this help, both within and across shifts, is an important source of complementarities, resonating with Itoh's (1991) theoretical argument that help efforts are cost efficient and give raise to team organization and compensation. To illustrate this point, consider an agent in a not-so-busy

shift who can choose to do nothing or prepare sandwiches for the colleague(s) in the next shift that might be more busy. Or, consider an agent who could invest her time, say, in cleaning the oven, but instead helps her colleague who is faced with a considerable queue of clients. In both cases, help is efficient, but in the kind of individualized work organization that is necessary for individual incentives, help may not be given.[2] Thus an individual bonus was out of the question.

Our research question is hence *not* whether team organization and the associated incentives increase efficiency, but rather, whether a team bonus, given an existing team technology, leads to economically significant efficiency gains, and through what channels these gains can be realized.

We find that the team bonus increased sales and customer visits in the treated shops by around 3%, equivalent to one third of the standard deviation, and wages by 2.3% on average, and up to 13% for some employees. The bonus was highly profitable for the firm, generating for each bonus dollar an extra $3.8 of sales, and $2.1 of operational profit. These effects are large for the retail sector in general, and particularly so for Germany, a country with high levels of managerial efficiency and intense product market competition.[3] Retail is one of the largest sectors in the world in terms of employment,[4] and, moreover, many firms in the global economy employ similar types of teamwork. This is the case, for instance, in catering or for the cabin personnel airlines that use a similar technology and are organized in teams. Hence, our results are likely to be widely applicable.

The treatment 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, the 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 was so profitable for the firm that the management decided to roll out the scheme to all shops after the experiment. We observe shop performance for another six months, from July to

---

[2] A respective theory paper is Auriol et al. (2002).

[3] According to Bloom and Van Reenen (2007) and Bloom et al. (2012), Germany has one of the highest levels of managerial efficiency in the world, second only to the U.S. This is particularly true for retail (Baily and Solow, 2001), a highly competitive sector, in large part because of the presence of two retail discounters, *Aldi* and *Lidl*, and low entry barriers (in contrast to, for instance, 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.

[4] In Germany, more than 3 million people (7% of the labor force) work in retail, and in the U.S. the figure is 14.9 million (10.2% of the labor force).

December 2014. During this period, the performance of treatment and control shops converged, an indicator for long-term stability of the treatment effect.

What are the mechanisms leading to the treatment effect? We show that it is unrelated to upselling (higher sales per customer visit), although this was, initially, an important element of the firm's strategy and reflected in sales guidelines and managerial activities. It appears instead that the team bonus provided a stimulus for improving operational efficiency. This interpretation is supported by the fact that the treatment effect on sales and customer visits are of a similar magnitude. Furthermore, we obtained data on prices of office space and apartments (courtesy to Germany's largest real estate platform *ImmobilienScout24*). We find that shops in zip codes with higher real estate prices have a higher treatment effect.[5] In shops in these areas it is likely that demand fluctuates more intensively (because office workers mainly shop at the beginning and end of the work day, and in their breaks). Moreover, the opportunity cost of waiting is higher for city office workers and customers with higher income.[6] As a result, improvements of operational efficiency that result in reduced waiting time will be more valuable in areas with high real estate prices. According to this interpretation, an incentive may stimulate to use its dispersed information in a way that is more fruitful for efficiency, and to invest team members' efforts accordingly.

We are not only interested in the efficiency effects of team incentives and what mechanism may be at hand. The heterogeneous treatment effects of the experiment also provide insights into the anatomy of teamwork, and lessons about the conditions under which team incentives are most likely to work well. A simple model helps organize the thoughts. Besides generating some predictions that are relatively straightforward (such as team effort will increase in the marginal returns to effort, and decrease with the marginal costs of effort), the model also predicts that it is the shops with a worse performance record that will react more strongly to the bonus, arguably because such shops have more slack and hence more wiggle room for improvements. These predictions are confirmed by the empirical results.[7]

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

---

[5] Running a regression with a coarser variable (size of towns) confirms this findings; the treatment effect is mainly driven by shops in large towns.

[6] We find no evidence for alternative explanations of the treatment effect, such as management input, "working smarter" (Burgess et al., 2010) through work shift reallocation, or increased friendliness (which we measured by a mystery shopping tour).

[7] In particular the last finding demonstrates how important the design of a compensation scheme is 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 the most effort.

bonus constant, one might expect that larger team size would reduce the treatment effect, because each member would receive a smaller bonus. However, our model shows that increasing team size can reduce or increase the treatment effect depending on the production technology and the curvature of the cost of effort function.[8] The data seem to indicate that, if anything, the treatment effects are stronger in larger shops.

The model also predicts the treatment effect to increase with the number of incentivized workers in a shop. In our study firm, not all workers were incentivized: the so-called "mini-jobbers", who earned up to €450 per month tax-free and made up around 30% of the headcount, had to be excluded from the bonus for tax reasons. This institutional specificity of Germany provides a source of exogenous variation in the share of non-incentivized workers in a team given the size of the bonus. We find that the treatment effect drops rapidly in shops with a higher proportion of mini-job workers. This finding is an indicator for effort complementarities between team members. Above a certain level of work hours provided by un-incentivized workers, the treatment effect drops to zero. This is not only evidence for the power of incentives and the importance of complementarities, it also shows that there are limits to peer pressure. In contrast to the theoretical analysis of peer pressure by Kandel and Lazear (1992) and the empirical work of Mas and Morretti (2009), the incentivized team members in our study firm seem to have failed in motivating their (un-incentivized) team mates.

Several implications of interest for researchers and practitioners alike follow from our findings. In particular, treatment effect heterogeneity may be taken as guidance for the applicability of team incentives across and within firms. A team bonus should work better for firms with a younger workforce, and in situations in which team members can affect the measured outcome substantially. Similarly, team incentives should work better in more urban and currently underperforming teams. While treating different shops differently may be cost-minimizing, the results of our field experiment indicate that unequal treatment *within* teams is detrimental for performance, and that peer pressure of the incentivized may not suffice to motivate the un-incentivized workers.

The results also allow us to draw some welfare implications. The welfare-enhancing effects of the bonus on both the firm's profit and employee wages must be set against possible costs in terms of higher stress levels incentive pay schemes may cause (Cadsby et al., 2007). We find, however, no adverse effects of the scheme on either employee turnover or job or life

---

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

satisfaction – at least during the period of observations. Thus, our bonus scheme is likely to be a win-win for the firm and the workers.

Our field experiment satisfies the methodological requirements of realism and randomization (List and Rasul, 2011). Employees work in an on-going firm, do not know that they are part of an experiment,[9] and carry out their normal day-to-day job, without any other intervention, except the introduction of a team bonus conditioned on pre-existing sales targets. In our setting, sorting is not an issue, because people are hired and assigned to shops centrally by the headquarters of the firm, and do not move between shops. This makes our study quite different from the existing studies on team, group and firm incentives discussed by Bloom and Van Reenen (2011) and Prendergast (1999). While those studies find evidence that incentives may increase efficiency, sorting constitutes a fundamental issue for identification (Prendergast, 1999). Hamilton et al. (2003) and Bandiera et al. (2013) have documented the substantial empirical relevance of sorting. Our randomization across units of the *same* firm also takes care of another identification issue discussed by Prendergast (1999): across firms, technology and profitability are likely to differ, and these differences are relevant for the decision in favor of teamwork (cf. Boning et al., 2007). Finally, many of the studies look at simultaneous changes in work organization (individual v. team) and incentives, while we keep teamwork fixed, and only change the compensation scheme.[10]

On a more general level, another contribution is to provide an insider perspective on the adoption of management practices (Ichniowski et al., 1997; Bandiera et al., 2011), complementing the literature on management practices across firms. Our study confirms Bloom and Van Reenen (2010), Syverson (2011) and Bloom et al. (2014) who argue that a main reasons why some firms adopt productivity-enhancing management practices and others do not, is product market competition. Internal firm politics is another factor influencing the adoption and success of new management practices. In particular, we argue that the same institutions that, in some instances, may create inertia and resistance to change, such as worker councils, may, in other instances, be conducive to reaching Pareto improvements between management and workers, because they are able to create high levels of commitment.

---

[9] Except for our partners in management and the workers' council (a German specificity whose importance is highlighted below), 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.

[10] Our paper focuses on the introduction of a monetary team incentive and is hence also quite different from recent field experiments that focus on the salience of existing incentive schemes (Englmaier et al., 2014), on relative performance evaluation between individuals (Barankay, 2012) and between teams (Delfgaauw et al., 2013), and lab experiments on incentives (Nalbantian and Schotter, 1997).

2. Background

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 that of the chains, but is sold at much lower prices, forcing the incumbent chains to rethink their business model.[11]

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 purchase, 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 their customers. However, with the changes in the market situation, the firm reacted to the challenges, and tried to develop new HR management practices aimed at improving the shopping experience.

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

Our study company has a well-defined hierarchical management structure, at the top of which are the general and district managers overseeing individual shops that are managed on a daily basis by shop supervisors. The shop supervisors ensure the efficient deployment of workers centrally allocated to their shops as well as compliance with technological and accounting procedures. However, they do not have a say in strategic matters (product mix, shop concept,

---

[11] The *New York Times* featured an article on the changes in the market for bread in Germany in its June 3, 2014 issue: *Caspar Oehlschlägel, who lives down the road from one of the oldest bakeries in Berlin, said that since his local supermarket started offering whole-meal bread baked in the store, it was all he ever bought. "Honestly, it's the best bread that I have ever had", Mr. Oehlschlägel said.*

prices, advertising campaigns, etc.) or personnel policies such as hiring, workforce composition, allocation, and incentives.

Before the changes in the market discussed above, our study firm offered incentives to its managers and shop supervisors but not to sales staff. For managers and supervisors there was a detailed system of key performance indicators (KPIs) 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 supervisors, 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 and supervisors in terms of their bonus. Sales performance is incentivized by a step-wise bonus the size of which depends on exceeding a predetermined and non-renegotiable sales target. The sales targets are determined at the end of the preceding year based on past sales and a correction for the general trend in sales (-2% in 2014).

Unlike shop supervisors, sales agents, who make up about 80% of the sales force, received fixed wages only (€9 – €11 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 mini-jobbers, workers who, in addition to receiving welfare benefits, earn up to €450 per month tax-free. The sales agents are predominantly unskilled, and employee turnover is high (see Table 1), 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 supervisors 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 the 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 proposed to our firm introducing a bonus payable to shop sales teams, including the shop supervisors, conditional on reaching or exceeding the sales targets. A surprised member of the management team said that "bonuses to sales staff were simply never on our agenda." Other members of the management team had considered a team bonus before but thought it would be ineffective because of free-riding. Some managers were also 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 and supervisors given the existing incentives, a sizable effect as we discuss later. Finally, it was unclear ex ante whether the additional wages that needed to be paid to those shops that would have reached their targets even in the absence of a bonus would be compensated for by the additional operating profits of higher sales.

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 was likely to be lower than €20,000 per month in the case where half of the shops were treated, and the maximum monthly shop (team) bonus was capped at €300. 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 increase in wage costs would reduce their own bonus if their personnel cost 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 was also in favor of the bonus, not only, because it was designed as a pure add-on payment, but also

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 our ability 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 the beginning of June, 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; 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 were satisfaction with the job context and overall satisfaction, both as constructed by Hackman and Oldham (1980) and translated into German by van Dick et al. (2001), and organizational commitment using the metrics developed by Allen and Meyer (1990). The June survey also collected some additional data we used for robustness checks. 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% in the first and 60% in the second wave of the survey. Although we randomized before the first survey, we find that the results of the survey are balanced between treatment and control group, a sign of the validity of our randomization procedure (explained below). However, comparing the attitude variables before and during the treatment, we found no treatment effect on either dimension of employee attitudes. We also found no significant interaction between baseline attitudes and the treatment effect on sales. In the remainder of the paper, we hence focus on the effect of the treatment on hard data, in particular, sales.

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 the employees 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 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.

At the same meeting, we trained district managers in how to explain the team bonus to the shop supervisors in the treatment group who would then relay our explanation to the employees in their shops. We also instructed the managers on 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 the event that they found out about the bonus scheme. We called the district managers every second week to inquire whether employees in the control group had heard about the team bonus. It turned out that questions about the team bonus were seldom asked. We will later 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, and wrote in the information leaflets sent to the treatment shops, that mini-jobbers had to be excluded from the bonus scheme because of tax reasons. According to German law, a mini-jobber who earns more than €450 in a month must pay taxes on their *entire* income, while below that income is tax-free. Therefore, giving a 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 to be shared between the part-time and full-time employees in the shop in proportion to their working hours in that month. The bonus increases by €50 for each percentage point above the target and is capped at €300 per month for exceeding the target by 4% or more. Hence, the team in a shop can make extra earnings of up to €900 in the treatment period of April to June 2014. We provided the employees with examples of what the sales increases would mean in terms of additional goods to be sold per day (for instance a 1% increase above the sales target for a mid-sized shop would be tantamount to selling per day ten additional rolls, two loafs of bread, two sandwiches and four cups of coffee).

FIGURE 1 ABOUT HERE

We realize that this bonus scheme may be criticized on theoretical grounds for being susceptible to the strategic behavior of employees around the bonus cutoffs (we label this as "gaming" in what follows). However, in 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 "threshold" bonuses are rather widely spread.[12] We do nevertheless address the possibility of gaming in Section 7.4.

3.3 Randomization and power of the experiment

We follow Barrios (2014) who shows that randomizing pairwise using the predicted outcome variable, in our case sales, minimizes the variance of the difference-in-difference treatment effect estimates. We use historic observations for 2012 and 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 three 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 the pre-treatment characteristics of our treatment and control shops. 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 per month worth of goods[13], employs seven people (four, in

---

[12] For example, the *World at Work 2012-2013* survey of incentive pay practices in approximately 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.

[13] One shop, located at a local transportation hub and assigned randomly to the treatment group, sold on average €118,000 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 per month with standard deviation of €10,885, 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.

full-time equivalents) 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%. Sales are quite variable, with location and size differences explaining 90% of the variance. There is also considerable sales 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 these factors we control for in our statistical analysis.

FIGURE 2 ABOUT HERE

Figure 2 displays the 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 4 million inhabitants. Shop locations vary in population size. However, almost all shops are placed on the premises of supermarkets, and therefore rely to some extent on customer traffic to and from grocery shopping.

4. Model

An illustrative model is helpful for generating hypotheses about the average treatment effect, the mechanism responsible, and treatment effect heterogeneity. Consider N agents working in a team creating output y that depends on total effort E, the productivity of team effort, a, and additive noise v with a probability distribution function $\phi(v)$ symmetric around, and centered at, zero:

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

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

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

The effort aggregation in (2) is flexible and can accommodate effort complementarity ($\rho < 1$) or substitutability ($\rho > 1$); when $p = 1$, the team's total effort is the sum of individual efforts. Note that we model the potential for complementarities in the benefits of effort, and not in the costs of effort as in Itoh (1991). Results would be quite similar, but at the expense of more involved modeling. As our model is mainly meant to be illustrative and complementarities may be present on both cost and benefit sides, we believe this to be a reasonable choice.

To model the incentive effect of the bonus scheme used in our study firm, we assume 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

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

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

The bonus is split evenly between the team members who independently decide on the individual level of effort e_i that maximizes their own payoff:

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

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 stems from some intrinsic motivation as in Holmström and Milgrom (1991, p. 33) or 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). Although under some constellations of the parameters (crucially, under effort complementarity, $\rho<1$), there may be multiple symmetric equilibria with high or low effort (Cooper and John, 1988), we focus on the high-effort equilibrium in which the team members choose an effort level e^*, $e_0 \leq e^* \leq e_{max}$, satisfying any one of the following sets of conditions:

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

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

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

or

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

or

$$e^* = e_{max} \text{ and } \frac{d\pi}{de_i}\bigg|_{e_i = e_{max}} \geq 0 \tag{7}$$

In words: there will be an 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. In the following Sections, we use comparative statics on the first-order conditions (5) for the interior solution $e_0 < e^* < e_{max}$ and present the respective empirical evidence.

5. Aggregate incentive effects and the mechanism

5.1. The effect of the bonus on effort

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

$$\frac{de^*}{dB} = -\frac{aN^{\frac{1-2\rho}{\rho}}\Phi'\left(aN^{\frac{1}{\rho}}e^* - y_0\right)}{\frac{d^2\pi}{de_i^2}} > 0 \tag{8}$$

TABLE 2 ABOUT HERE

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. Notice first that for the entire chain, sales and customer visits have gone down, reflecting the secular downward trend in the bakery business. However, significant differences between treatment and control shops are noticeable for sales and customer visits, suggesting a positive treatment effect. In fact, the difference-in-difference estimated effects on the log sales and customer visits are 3.3% and 2.8%, respectively, both significant at conventional levels. Other variables, in particular employee attitudes show no changes. Hence, we proceed with a more in-depth analysis of sales (this section) and customer visits (Section 7.1).

FIGURES 3 AND 4 ABOUT HERE

To visualize the treatment effect on sales, we plot post-treatment changes in sales rankings of shops within the treatment-control pairs that resulted from our assignment procedure. (Recall that it consisted in flipping a coin for the pair of shops with adjacent ranks in terms of predicted sales performance.) Figure 3 shows that only 18% of the shops in the control group improved their within-pair ranking after the treatment, while in the treatment group it is 37%. Additionally, Figure 4 plots the kernel density graphs of the year-on-year sales growth for treatment and control groups and shows a uniform shift in the treatment group's sales growth distribution to the right from the control group.

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

$$\ln(sales_{it}) = \beta * treatment_i * after_t + period_t + shop\ fixed\ effect_i \qquad (11)$$
$$+ controls_{it} + error_{it}$$

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.

TABLE 3 ABOUT HERE

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.[14] The average treatment effect is 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 for our baseline results, we allow for correlation between the treatment status and the baseline outcome, which, despite randomization, may occur in finite samples and cause "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).

Whatever specification we use, we obtain 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 for our relatively large sample.

---

[14] For each of these outliers we identified the reasons for the deviations. We found that there are two main reasons: temporary closures, for instance related to refurbishments, and sales drops because of refurbishments of roads or malls. Outliers occur both in treatment and control group, and in treatment and baseline periods.

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, we find it to be 2.9% in April, 3.7% in May, and 2.9% in June 2014, a steady effect without noticeable abatement.

5.2 The effect of the bonus on profitability

Comparing the gains of the bonus with the costs of its implementation, we can first note that the average treatment effect on sales of 3% implies an extra €820 (=[exp(0.03)-1]*€27,000) worth of sales per month in the average shop, or €238,620 (=€820*3 months*97 shops) in all treatment shops over the treatment period. Given the historic share of value added in sales of 0.56, the implied operational profit gain is €133,627.

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 or 3.9% of the average recipient's quarterly earnings. The total team bonus payments in April to June 2014 amounted to €35,150, or 2.3% of the total labor costs in the treatment shops. There was a knock-on effect on shop supervisor bonuses: €240 per treatment shop per quarter (= difference-in-difference estimate of the treatment effect on the shop supervisor bonus), adding an extra €23,280 for all 97 treatment shops. In addition, district and top manager bonuses increased by an estimated €4,500. There were also one-off costs associated with the implementation of the bonus scheme: printing and delivering materials, administrative support (bonus calculations and communications) and the costs of managers' time required to implement the scheme, which we estimate at €25,000.[15] The total costs add up to €87,930.

The benefit from the scheme net of the costs is €45,700 for the treatment shops in the treatment period. Projecting our calculations to the time past July 2014, when the scheme was rolled out to all 193 shops, the implied net gain becomes €140,000 per quarter for the entire chain. Our calculations imply that each dollar spent on the bonus brings $3.8 of extra sales, or $2.1 of extra operational profit.[16] In sum, our scheme is a viable "investment in people" project and a win-win for the firm *and* its workers.[17]

---

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

[16] We were not paid for our advice, but one could consider the costs of our time up to implementation (i.e., not related to the research). Evaluated at the cost of a major consulting firm, the break-even of the project would be reached in less than a quarter.

[17] Another project the firm undertook was to invest in a thematic redesign of 31 selected shops. The profitability of this project is far smaller than that of the bonus scheme. Estimating the sales response in the 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

5.3 The mechanism

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 a combination of both. We dissect our estimated treatment effect along these margins, showing that the treatment effect has occurred along the extensive margin and that much of it owes itself to improved operational efficiency.

Starting with the intensive margin, we find a precisely estimated but close to zero treatment effect of 0.4% 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. They were unaware about whether or not the shop was in the treatment group. They took 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.

Turning to the extensive margin, we find 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.[18] In principle, within the existing operational constraints (opening hours, product ordering rules, standards of service, etc),[19] incentivized shop teams could increase the number of customer visits by *working smarter* (that is, reallocating labor hours across work shifts to meet the fluctuating demand), or by *working faster*. Working smarter is unlikely to be a major driver of the treatment effect because reallocating labor hours takes time – at least a month under the company rules; the treatment

€150,000 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, the average return on investment over a ten-year horizon would be less than 0.6% per year.

[18] One could argue that the gain in sales and customer visit could have come from personal contacts of the now incentivized shop agents. Against this hypothesis are the facts that i) the average increase in customer visits is by 250 a month, improbably large to be supported by the personal connections of only five incentivized employees in an average shop; and ii) the treatment effect does not depend on the median employee home to work distance, which we used as a proxy for the co-location of employees' social networks.

[19] In the next subsection, we argue why these constraints cannot be relaxed because of hierarchical and legal reasons.

effect, however, is stable in all months during the treatment period; 2.9% in April and June, and 3.7% in May.

There is support for the working faster story, though. Working faster makes it possible to serve more customers per unit of time and to have more time to serve customers by freeing time up from behind-the-counter tasks (such as cleaning or delivery). There are, moreover, numerous ways in which shop assistants can allocate their effort across and within shifts in order to increase overall sales. After the experiment, we learned from interviews with the shop assistants that within shifts and in peak times, it is crucial that teammates help each other in dealing with the customer. If one employee both operates the cashier and passes the client the good they paid for, the queue will be managed slower than when one employee operates the cashier and another passes the goods. Across shifts, employees can help deal with peak times by baking fresh bread at the end of a shift with low demand, such that the next shift can focus on the customer. Similarly, in low-demand shifts, employees can prepare sandwiches, clean the coffee machine etc., rather than doing nothing. This effort can help avoid the loss of customers because queuing times during peak hours are too long.[20]

An important element to finding support for this story in the data is that for some, but not all, shops there exist such peak times and that the opportunity costs of waiting differ across shops, and so does the wiggle room for working faster. We cannot identify peak times within a day or a week, but we can distinguish treatment effects by shop location. This should affect the magnitude of effort response to a given incentive by influencing the marginal product of effort.

Before going to the data, consider the comparative statics on the first-order conditions for the optimal effort level under the team bonus. Effort increases with the productivity parameter a:

$$\frac{de^*}{da} = -BN^{\frac{1-2\rho}{\rho}} \cdot \frac{\Phi'\left(aN^{\frac{1}{\rho}}e^* - y_0\right) + N^{\frac{1}{\rho}}e \cdot a\, \Phi''\left(aN^{\frac{1}{\rho}}e^* - y_0\right)}{\frac{d^2\pi}{de_i^2}} > 0 \qquad (9)$$

The expression in the numerator of $\frac{de^*}{da}$ is positive when

$$\left|\Phi''\left(aN^{\frac{1}{\rho}}e^* - y_0\right)\right| \ll \Phi'\left(aN^{\frac{1}{\rho}}e^* - y_0\right).$$

---

[20] Note that the density of bakeries in the market of our chain is large, and that most bakeries are located in supermarkets offering close substitutes to the customers.

This is the case when the team's output, $aN^{\frac{1}{\rho}}e^*$, is close to performance target y_0 and N is not too large.[21]

Empirically, we should expect that 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 agents may succeed in catering to both these groups by saving their time in the queue. In contrast, in smaller locations, there is less competition, and hence less choice for customers. The demand of these more regular shoppers 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. Looking at the treatment effect by shop location in the preferred difference-in-difference specification (1), we found that the treatment effect is largest, at 5.5%, in shops located in big towns (>60,000 inhabitants), going down to 3.8% in midsize towns, and zero in villages.

To investigate this channel further, we investigated the finer grained data from *ImmobilenScout24*, the leading real estate platform in Germany. We use zip code level residential and commercial property data to device proxies for the incidence of peak times and opportunity cost of waiting in the queue. The data include prices per square meter of residential and commercial properties in a total of 136 zip codes in which our shops are located. We hypothesize that the incidence of peak times is more likely in areas with a higher presence of commercial properties, and the opportunity costs of waiting in the queue are higher in more expensive areas. Accordingly, we interact our treatment effect with the average rental price for commercial and private real estate.

TABLE 4 ABOUT HERE.

The estimation results, reported in Table 4, demonstrate that the treatment effect does increase with local property prices, both residential and commercial. A one-standard-deviation increase in rental prices drives the treatment effect up by about 0.03 from the average of 0.03. The high correlation between office and residential rental prices (0.8) suggests that these two measures reflect the overall attractiveness of the area rather than its business vs. residential flair. We interpret the higher treatment effect in pricier zip codes as strong evidence for working faster as the mechanism through which team bonus resulted in higher sales. It is more expensive areas where time is more precious and hence the reduction in queuing through working faster is more likely to prevent losing customers to competitors.

---

[21] 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.

5.4 Excluding other mechanisms

There are several alternative mechanisms that can be rejected. 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. Not only are they forced by their rental agreements to exactly follow their host's opening hours, but for most shops it is physically impossible to remain open when the supermarket or mall on whose premises they operate is closing.[22] Second, one could imagine that the shops would order more products from the central warehouse to satisfy customer demand – at the cost of higher share of unsold goods. However, the automated ordering system gives little room for flexibility in orders. Moreover, there is no treatment effect on the share of unsold goods, either as can be seen in Table 2, Panel A. Third, 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. Their evaluations, either with or without mystery shopper fixed effects, are uncorrelated with sales, which goes against the hypothesis that friendlier customer service is behind the observed treatment effect.

We also argue that the management input of shop supervisors is unlikely to be a major channel behind the treatment effect, for the following reasons. First, the interaction of the treatment with worker characteristics, most notably, the share of un-incentivized mini-jobbers, implies that shop supervisors' ability to affect sales is constrained – to such a big extent that the treatment effect is zero for some shop teams. Second, the marginal increase in the shop supervisor bonus as a result of team bonus – €80 per month on average – is rather small compared to their individual bonus (up to €500 per month). Despite the sharp incentives available to shop supervisors, sales targets were reached only in about 35% of the time before the team bonus – another argument suggesting that shop supervisors' influence on sales has its limits.

Third, steeper shop supervisor incentives with respect to sales, implemented in January 2014, did not produce a significant effect on sales. Before January 2014, shop supervisor bonus depended on sales, personnel costs and the mystery shopper score. Then, influenced by our findings that mystery shopping scores were too subjective to be a valuable performance indicator (the topic of another project of ours), management decided to remove

---

[22] There are around 30% of shops in which this is not the case, because they have a separate entrance. These shops could (potentially) extend their hours. Removing these shops does not change our results. Besides, assuming that the entire sales gain of 3% was achieved by working longer rather than working faster implies that an average shop should have been open for 30 additional minutes per day, which is impossible under the current regulations.

the mystery shopping criterion from the rules determining the bonus. The increased importance of sales for the supervisor bonus should have affected the supervisors' effort directed to sales. However, backing out the implied effect of the change in shop supervisor compensation under the assumption of a constant trend in sales, we find the effect of the change in supervisor incentives on sales to be 0.5%, and insignificant.

Finally, we find that the team bonus effect on sales is not affected by whatever proxy we use for shop supervisors' management input. The proxies we used are: shop supervisor monthly working hours, tenure, average bonus he or she received between January 2012 and March 2014, leadership score from the employee survey in June 2014 (Carless et al., 2000), and the linear combination of the above four proxies with weights estimated from the production function regression of shop sales on shop, worker and supervisor characteristics. These proxies are correlated with shop sales before the treatment (R^2=0.13, F-stat=127). None of our shop supervisor 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 supervisor input significantly affects the magnitude of the effect of team incentives on sales. Consistent with this interpretation is the fact (recall section 2.2) that shop supervisors cannot pull many levers affecting shop performance; indeed, neither workforce characteristics nor turnover changed in response to the bonus.

6. Treatment 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). We identify the following heterogeneous treatment effects (the proofs are in Appendix II):

1. Individual effort decreases with the difficulty of the costs of effort parameter b:

$$\frac{de^*}{db} = -\frac{-b}{\frac{d^2\pi}{de_i^2}} < 0 \qquad (10)$$

2. Individual effort decreases with team size N if effort complementarities are not too strong ($\rho > 1/2$). 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.

3. Team effort decreases with the share of non-incentivized members in the team.

4. The effort under the bonus will depend on the frequency of reaching the targets in the past, without the bonus. The effort response to the bonus of more successful teams 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.

Let us briefly discuss how these predictions match into observable heterogeneity. Prediction 1 is on the cost of effort, and we will use employees' age as a measure, arguing that young employees should have lower costs of effort. The next prediction is on team size, which we measure directly. Prediction 3 is directly measurable by the proportion of total work in a given shop carried out by mini-jobbers. Finally, Prediction 4 is also measurable, as we have data on 27 months of shop performance before the introduction of our experiment.

We report the results of testing these predictions in Table 5. Our treatment and control groups are balanced in all the characteristics we analyze below.

TABLE 5 ABOUT HERE

According to Prediction 1, we expect the treatment effect to be larger for a younger workforce, since younger workers might have lower effort costs. Besides, there may be an element of resistance to change, which can be assumed to be weaker among younger workers. Panel A in Table 5 reports treatment effects in the shops below and above the median workforce age based on our preferred difference-in-difference specification. 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: the treatment effect interacted with age and tenure separately as well as jointly produces a significant interaction with age but not with tenure.

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 Prediction 2, 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. Panel B in Table 5 shows that the treatment effect is larger in bigger shops. The observed differences in the treatment effect are not due to bigger shops being located in bigger towns.

Prediction 3 is on the impact of the share of mini-jobbers work in the total work hours delivered by a shop team. More 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.[23] To accommodate the latter, 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 C of Table 5. We find that the treatment effect goes down with the share of mini-jobbers. The abrupt drop in the treatment effect to zero past the second 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.

According to Prediction 4, we expect the treatment effect to vary with the past performance around the sales target. Historic success in 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) 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 level.

Panel D of Table 5 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 D1); and ii) the frequency of a shop achieving its target in the pre-treatment period (Panel D2). 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.

7. Robustness checks

We consider four scenarios in which our estimation results may have been driven by factors other than the team bonus treatment. First, despite randomization succeeding at balancing the control and treatment samples before treatment, the treatment group shops may have been on

---

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

a different trend from control group before, during or after treatment. Second, sales assistants 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, the bonus treatment 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 label this possibility as contamination). Finally, 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 confront these scenarios with data; there is no support for either of them.

7.1 Differences in sales trends and "post-treatment treatment effect"

Figure 5 plots log shop average monthly sales by treatment and control group, showing no significant between-group differences in any month during the pre-treatment period. (The formal statistical test confirms this observation.) Hence, there are no pre-treatment differences in sales trends either. There is a stable treatment/control difference in sales during the treatment period. In the post-treatment period starting from July 1st 2014, when the bonus scheme was rolled out to all shops, and ending in December 2014, sales become more varied in both groups. There seem to be some treatment v. control differences in July, September and November 2014, but only the July difference is statistically significant. While the roll out began in the end of the month of June, it took district managers a while to inform all their shops about it. Partly this was owing to the fact of a firing and reassignment of district managers to shops, resulting in larger span of controls and higher workloads for most of them.

We nonetheless carried out some further analyses to exclude the existence of a spurious post-treatment "treatment effect", which could cast doubt on attributing our main result to the effect of the team bonus.

FIGURES 5 AND 6 ABOUT HERE

First, repeating our treatment effect analysis in section 5.1 for the period July to December 2014, we find no difference in within-pair ranking changes between the (former) treatment and control group shops (Figure 6). This result is in stark contrast to the significant treatment/control differences in the frequency of within-pair rank improvement during the treatment period of April-June 2014 (Figure 3).

FIGURE 7 ABOUT HERE.

Second, Figure 7 plots treatment effect estimates and their 95% confidence intervals separately for each month since April 2014 from an extended version of our baseline regression specification (11):

$$\ln(sales_{it}) = \sum_{k=1}^{2} \gamma_k \, ta_{it+k} + \sum_{k=0}^{8} \beta_k \, ta_{it-k} + period_t + shop\ fixed\ effect_i + controls_{it}$$
$$+ error_{it}$$

where $ta_{it} = 1$ if a shop i is in the treatment group and the time period is April 2014, and 0 otherwise. The treatment effect k months after the start of the treatment is thus β_k. A significant post-treatment "treatment effect" (PTTE) could jeopardize our story; the PTTE is, however, small and insignificant in all months except July 2014, the month during which the roll out took place.[24] The average PTTE in July-December 2014 is just 1% and insignificant (p-value=0.467), whereas its difference from the average treatment effect of 3% is statistically significant (p-value<0.05). Finally, we include leads of the start of the treatment dummy ta_{it} with coefficients γ_k to control for possible anticipation effects. There were none.

In sum, with no significant differences in sales between the treatment and control groups outside the treatment period, and with no detectable treatment effect once the team bonus was rolled out, the introduction of team bonus is the only plausible cause behind the treatment effect we have found.

7.2 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 (2015), which also checked for the 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, we were informed by management that 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.

---

[24] The roll out of the team bonus scheme in that month coincided with a major reassignment of area managers undertaken by our study company without our involvement. The reassignment caused some district managers having to deal with many new shops – 27 in one case – which may have prevented communicating the bonus scheme to the former control shops in good time.

Finally, our data are not supportive of Hawthorne effect either. Using the rollout episode, we construct the Hawthorne-free treatment effect (HFTE) and compare it with the baseline treatment effect (BTE) which could have been influenced by Hawthorne. BTE being larger than HFTE would indicate the presence of Hawthorne effect. We estimate HFTE as the difference between the control group log average sales in August to December 2014 (10.14), and what they would have been in the absence of the bonus scheme under the constant trend assumption (10.10).[25] The estimated HFTE effect of 4% is Hawthorne-free because the scheme had already been rolled out to all shops in August 2014. Comparing HFTE of 4% with BTE of 3%, we conclude that our baseline estimates are not prone to Hawthorne effect.

While neither the institutional context nor data provide any evidence consistent with the Hawthorne effect on worker effort, the possibility of being so affected still exists 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 June 2014 employee survey we learn that there is no difference in the frequency of district manager visits between the treatment and control shops (four to five visits per month on average in both groups).

7.3 Contamination

We have taken great effort to prevent contamination between the treated and non-treated shops in our experimental design. Indeed, as Bandiera et al. (2011) have argued, it is important to isolate treatment and control groups either geographically or in terms of the information available, or both. While we decided, on the grounds of randomization, against separating the control and treatment shops geographically, we took steps to separate them informationally and, should that fail, to detect information leakages, and mitigate the consequences. Hence, we did not let the workers in the control group know that 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,

---

[25] Recall the same method used in section 5.4.

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 almost 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 the interaction coefficient is insignificant (p-value of 0.5). Summarizing, all of our contamination tests fail to provide evidence for contamination.

7.4 Gaming

As mentioned before, 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, and another one by €8) – 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 8 ABOUT HERE.

Figure 8 offers a more systematic perspective on the potential 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% 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 of 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 7.6% on the frequency of excess sales above 4.5%. Summing up, there is little if any evidence for gaming, and even if there were some gaming it would fail to explain the magnitude of the treatment effect we have found.

8. Discussion

8.1 Getting it done: The political economy of 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 the performance effects of management practices, 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 points. A lack of awareness ("[Monetary incentives to sales staff] were simply never on our agenda") 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 in working 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 convinced the 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 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 bonuses for sales staff would certainly imply higher personnel costs, whereas the sales benefits were not clear at the beginning; hence the initial skepticism of middle management, whose bonuses depended on both sales and personnel costs. It took the clearly articulated commitment of the CEO to earmark a separate budget for the team bonus to overcome this resistance. Second, while some employees stood to gain from a team bonus, others would lose. 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.

8.2 Practical aspects of management practice implementation

List (2011) presents a guide for field experimenters, which inspired elements of our own experimental design. Our experiment was guided by economic theory to inform our treatment and interpret the findings, and we spent substantial effort on randomizing and measuring the statistical power of our experiment according to the current best scientific practice; we managed to find a champion for our cause in the top management; and addressed organizational complications to our experiment.

The most instructive experience we made was in dealing with organizational resistance to new practice implementation (tip 6 in List, 2011). To remind, the sources of initial organizational resistance to our "pilot" were owing to conflicting incentive schemes for managers (existing) and sales staff (new), and to extra burden on the HR personnel. We addressed these two causes for resistance by allocating 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, tip 11) 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 in November 2013 (see Section 5.4), 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. Here again, our experience with the workers council seems instructive: it suggests that institutions that one may expect to be an obstacle to change and experimentation, when convinced, will help the experimenter by boosting trust and legitimacy.

9. Concluding remarks

Teams are a ubiquitous feature of modern production, and so are monetary incentives. While knowledge about the effectiveness of individual incentives is broad and deep, much less is known about team incentives. Endogeneity, complementarities and employee sorting make causally interpretable evidence about the effectiveness of team incentives hard to obtain. Our large-scale natural field experiment involving 193 shops and 1,300 employees of a bakery chain in Germany provides such evidence. The estimated treatment effect is around 3%, or one third of the sales' standard deviation. There is substantial heterogeneity, with the treatment effect being largest in big towns, and in shops with a 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 achieved 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 N., 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, "Social Connections and Incentives in the Workplace: Evidence from Personnel Data," *Econometrica*, 77 (4) (2009), 1047-1094.

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 Firm Level Experiment," *Journal of the European Economic Association*, 11 (5) (2013), 1079-1114.

Barankay, Iwan (2012), "Rank Incentives: Evidence from a Randomized Workplace Experiment," *Working Paper*.

Barrios, Thomas, "Optimal Stratification in Randomized Experiments," Mimeo, Harvard University (2014).

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Difference-in-Differences 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.

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*, 130 (1) (2015), 165-218.

Bloom, Nicholas, and John Van Reenen, "Measuring and Explaining Management Practices cross 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 (547) (2010), 968-989.

Burks, Stephen, Jeffrey Carpenter, and Lorenz Goette, "Performance Pay and Worker Cooperation: Evidence from an Artefactual Field Experiment," *Journal of Economic Behavior & Organization*, 70 (3) (2009), 458-469.

Cadsby, C. Bram, Fei Song, and Francis Tapon, "Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation," *Academy of Management Journal*, 50 (2) (2007), 387-405.

Cameron, A. Colin, and Douglas L. Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50 (2) (2015), 317 – 372.

Carless, Sally A., Alexander J. Wearing, and Leon Mann, "A Short Measure of Transformational Leadership," *Journal of Business and Psychology,* 14 (3) (2000), 389-405.

Cooper, Russel, and Andrew John, "Coordinating Coordination Failures in Keynesian Models," *Quarterly Journal of Economics*, 103 (3) (1988), 441-463.

Courty, Pascal, and Gerald Marschke, "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.

Friebel, Guido, and Wendelin Schnedler, "Team Governance: Empowerment or Hierarchical Control," *Journal of Economic Behavior & Organization*, 78 (1-2) (2011), 1-13.

Friebel, Guido and Levent Yilmaz (2015), "Flexibility, Specialization and Individual Productivity: Evidence from Call Center Data", *Working Paper*, Frankfurt.

Hackman, J. Richard, and Greg R. Oldham, *Work Redesign*, Reading: Addison-Wesley, (1980).

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.

Ichniowski, Casey, Kathryn L. Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 87 (3) (1997), 291-313.

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- 145.

Mohnen, Alwine, Kathrin Pokorny, and Dirk Sliwka, "Transparency, Inequity Aversion, and the Dynamics of Peer Pressure in Teams: Theory and Evidence," *Journal of Labor Economics*, 26 (4) (2008), 693-720.

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

Prendergast, Canice, "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37 (1) (1999), 7-63.

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.

van Dick, Rolf, Christiane Schnitger, Carla Schwartzmann-Buchelt and Ulrich Wagner, "Der Job Diagnostic Survey im Bildungsbereich," *Zeitschrift für Arbeits- und Organisationspsychologie*, 45 (2) (2001), 74-92.

von Siemens, Ferdinand, and Michael Kosfeld, "Team Production in Competitive Labor Markets with Adverse Selection," *European Economic Review*, 68 (2014), 181-198.

World at Work 2012-2013, available at http://www.worldatwork.org/adimLink?id=59595

Figure 1: The team bonus

monthly bonus

300 €

250 €

200 €

150 €

100 €

0

% above sales target

Target   + 1 %   +2%   +3%   +4%
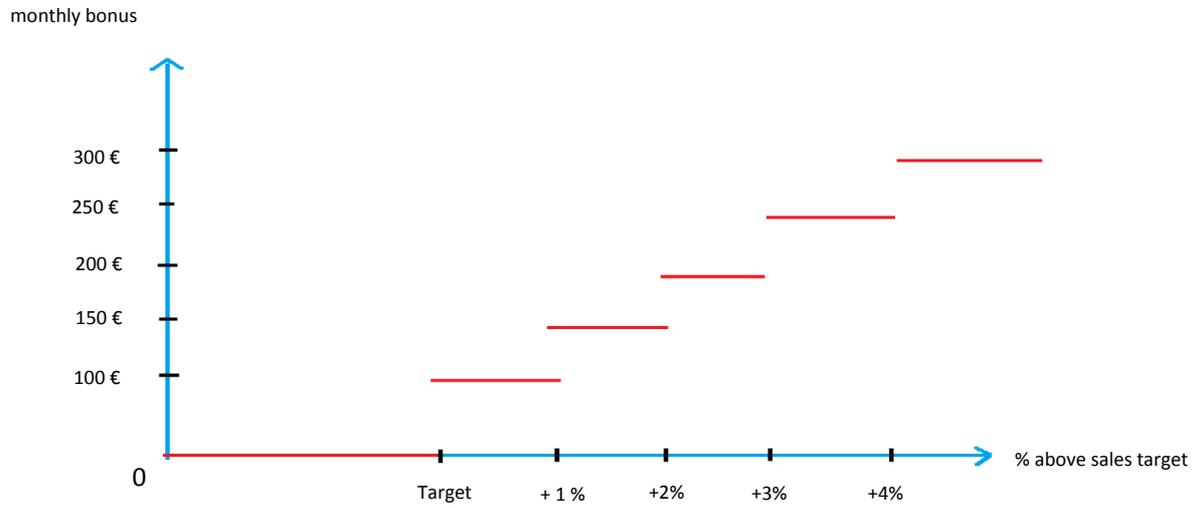
Figure 2: A map of shops by control (black) and treatment group (white)

© www.landkarten-erstellung.de
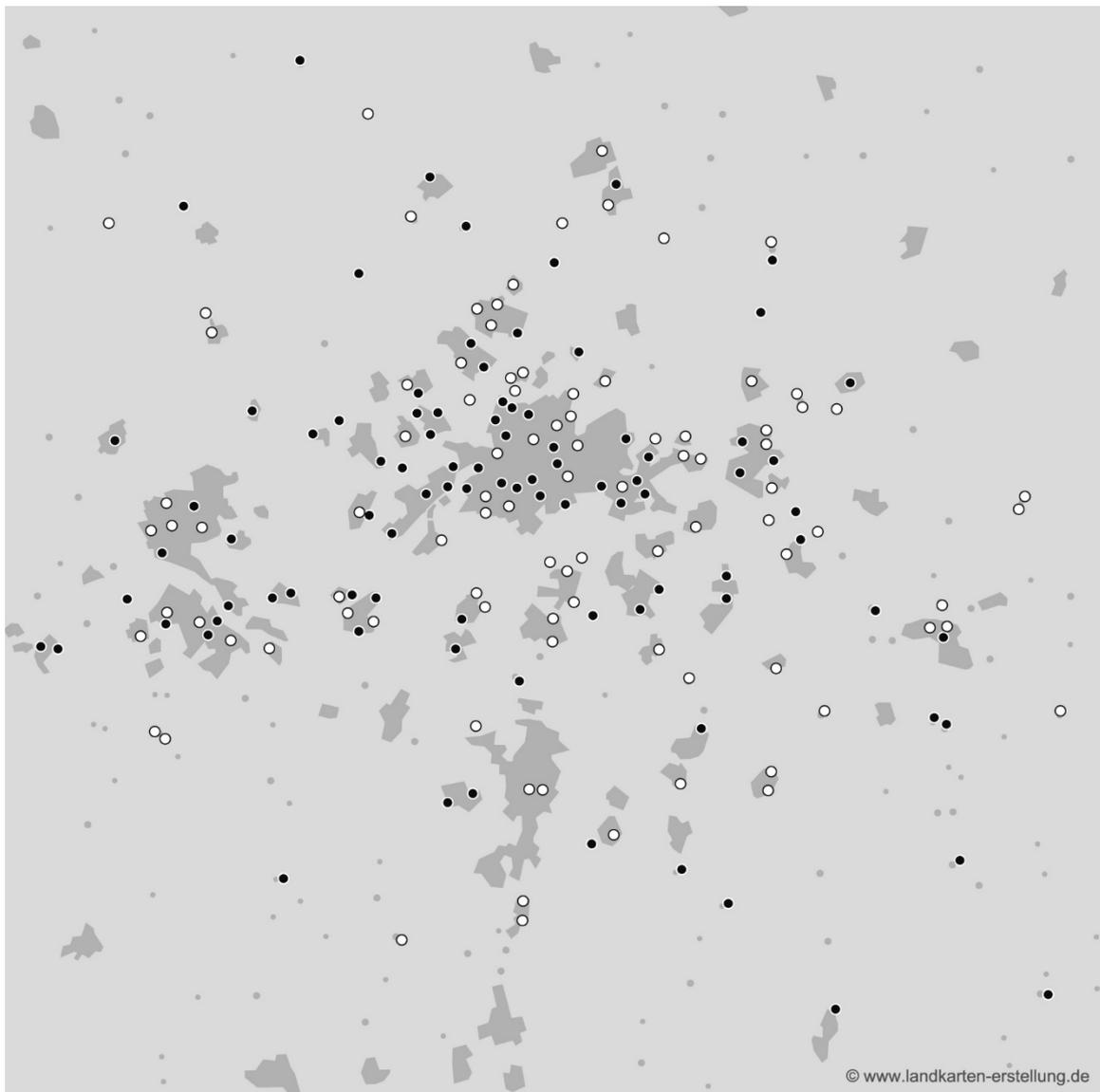
Figure 3: Improvement in within-pair sales ranking by group, April-June 2014
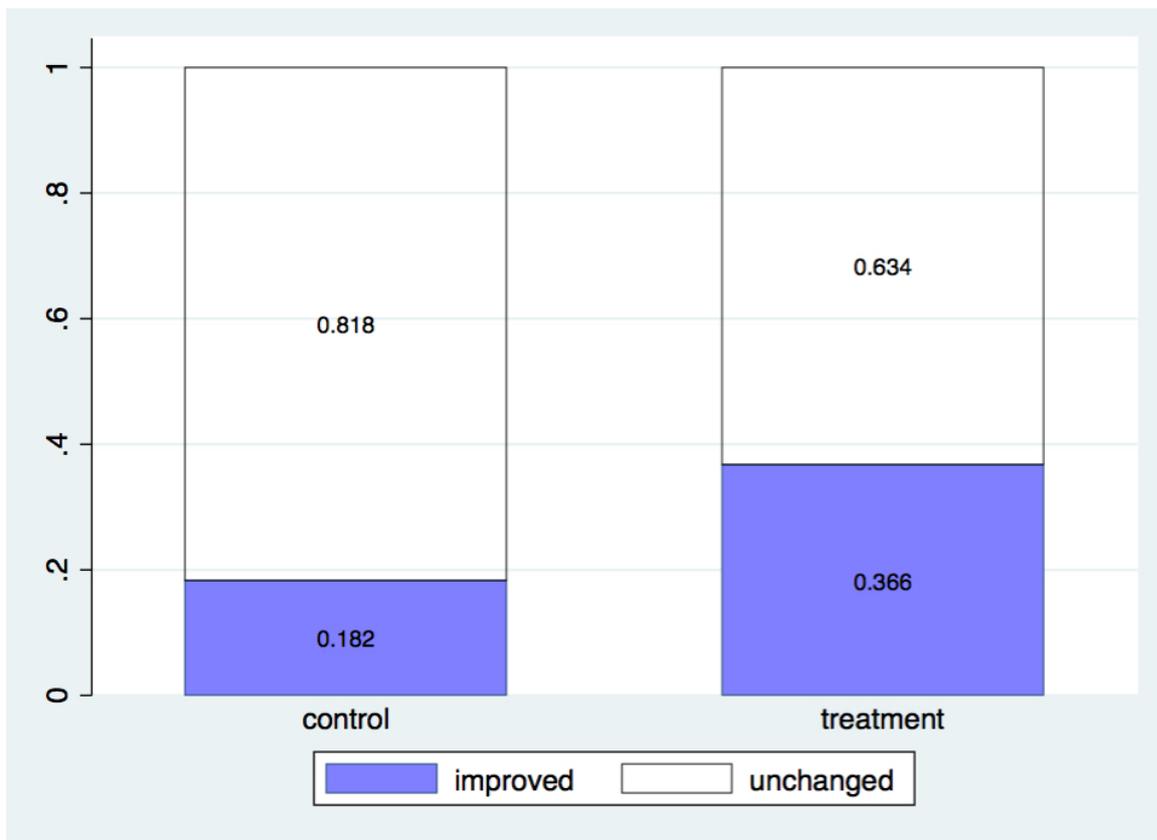
Figure 4: Kernel distribution of the year-on-year sales growth

kernel = epanechnikov, bandwidth = 0.0100

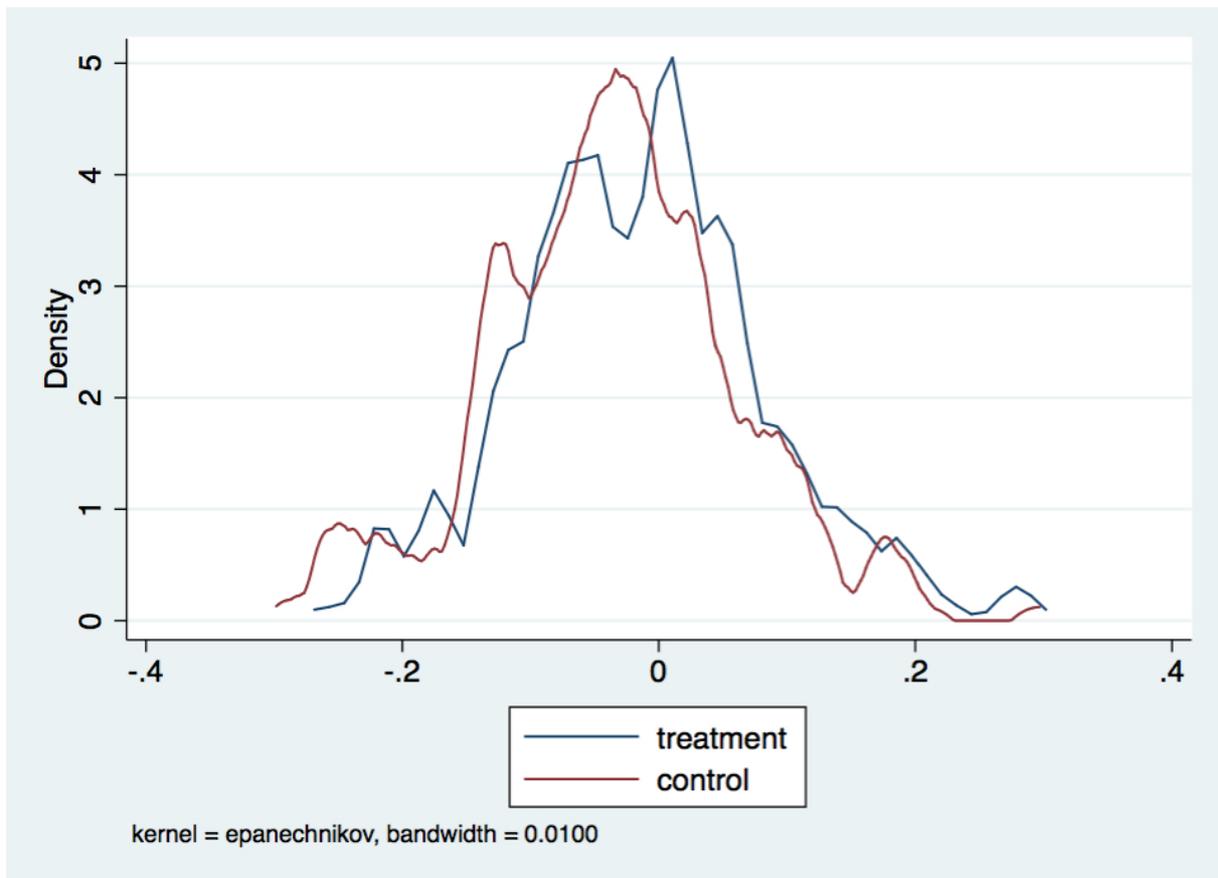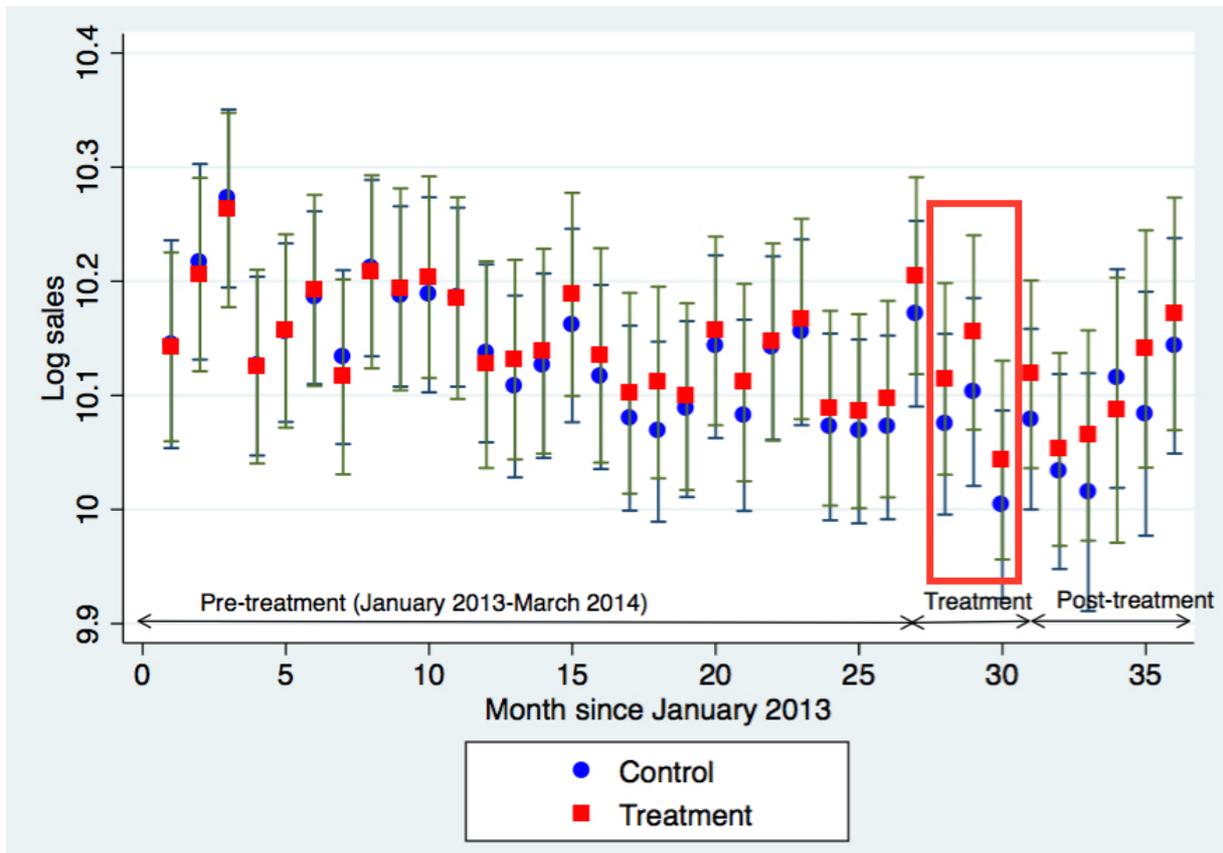**Figure 5: Log average monthly shop sales by group**

Note: vertical bars measure 95% confidence intervals.

Figure 6: Improvement in within-pair sales ranking by group, July-December 2014

Figure 7: Treatment effect by month of 2014

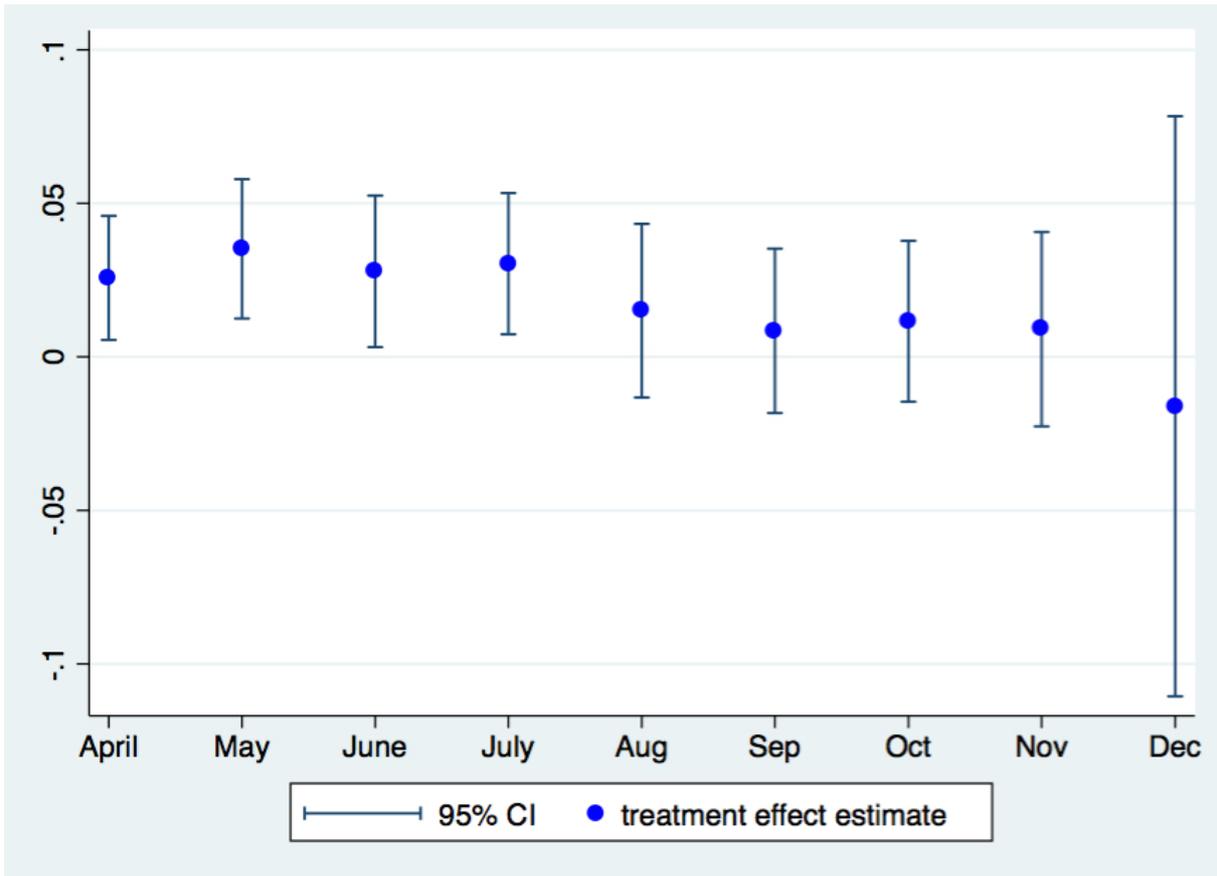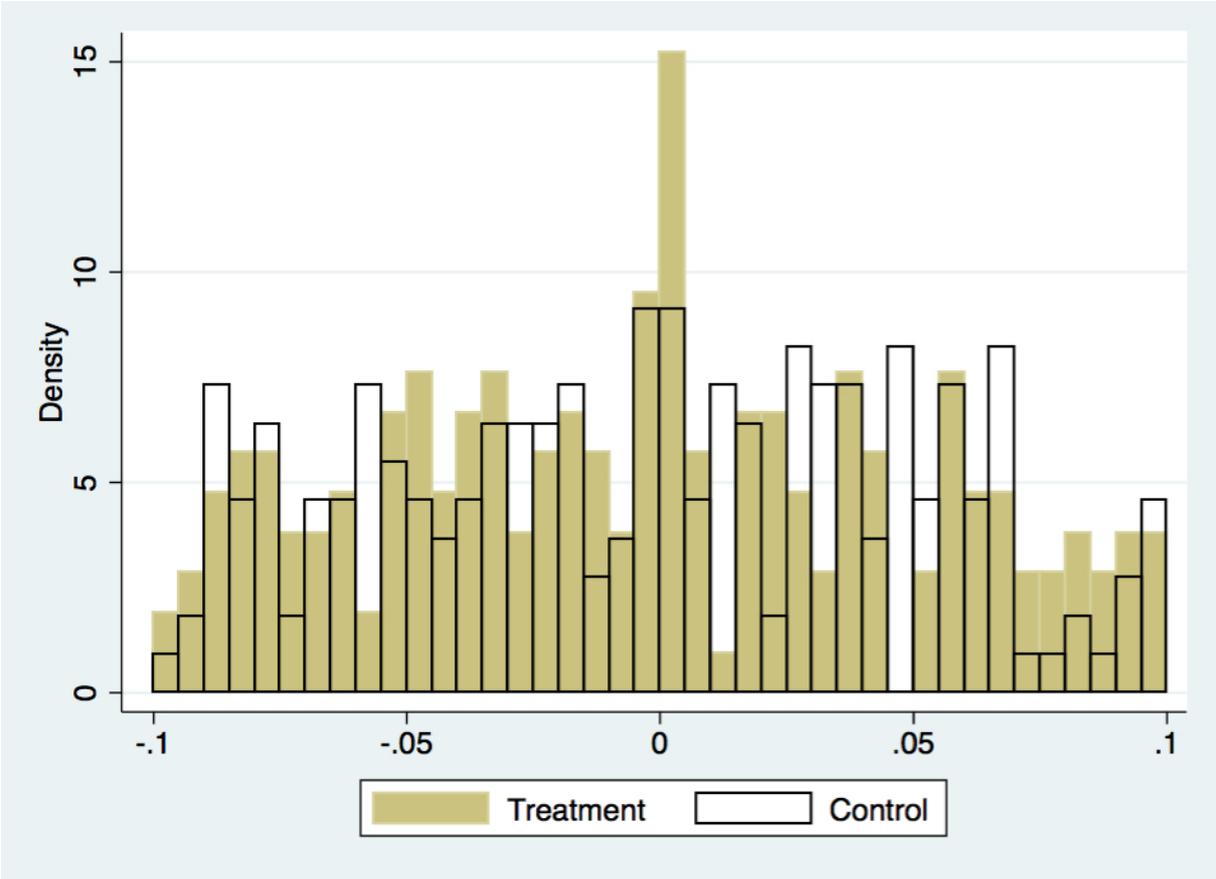**Figure 8: Percentage deviation of sales from the target in the treatment period**

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

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

	Panel A: Quantitative performance indicators		
	Control (n = 96)	Treatment (n = 97)	t-test p-value
Mean monthly sales (SD)	27,453 (11,481)	28,194[a] (14,542)	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)	10,028 (3,921)	10,131 (4,018)	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 contract	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

[a] One shop sold on average €118,000 worth of goods per month in the pre-treatment period. Excluding this shops, the average pre-treatment sales in the treatment group are €27,176 per month (standard deviation: €10,885).

Note: 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 work satisfaction and overall satisfaction scores constructed by Hackman and Oldham (1980) and translated into German by van Dick et al. (2001) and commitment 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)	25,376 (10,708)	26,995 (15,036)	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)	9,115 (3,582)	9,465 (3,790)	0.062
Mean monthly quit rate (SD)	1.4% (4.9%)	1.7% (5.6%)	0.336
Frequency of achieving the sales target	44.8%	49.1%	0.442
	Panel B: Qualitative performance indicators		
Mean mystery shopping score	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

Note: 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	4916	4904	386	193	577	561	577

Note: 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 by shop zip code property market characteristics

	(1)	(2)	(3)
Treatment effect at mean	0.035	0.031	0.035
	(.013)	(.013)	(.013)
TE * Average per sq. meter price: Commercial	0.036		
	(.019)		
TE * Average per sq. meter price: Residential		0.031	
		(.012)	
Weighted average by sq meters			0.031
			(.014)
Observations	4739	4889	4739

Note: The regression specification is the same as spec. 1 in Table 3. Standard errors are clustered by shop. All the variables interacting with the treatment effect are standardized to have mean zero and standard deviation 1. Thus, for example, the coefficient 0.036 on the interaction between the treatment effect and the average rental price of commercial property in spec. 1 implies that a one standard deviation increase in commercial rental price is associated with a 0.036 higher treatment effect.

Table 5: Treatment effect heterogeneity

Panel A: Treatment effect by shop-average employee age	
Above median	Below median
0.001	0.061
(.017)	(.019)

Panel B: Treatment effect by quartile of shop size (number of workers)			
Quartile 1	Quartile 2	Quartile 3	Quartile 4
0.001	0.022	0.041	0.059
(.024)	(.022)	(.027)	(.025)

Panel C: Treatment effect by the average share of mini-job employees			
Quartile 1	Quartile 2	Quartile 3	Quartile 4
0.071	0.050	0.003	-0.003
(.033)	(.026)	(.019)	(.021)

Panel D: Treatment effect by pre-treatment deviation of sales targets			
D1: Distance measure: pre-treatment average sales/target difference			
Quartile 1	Quartile 2	Quartile 3	Quartile 4
0.046	0.036	0.047	0.003
(.026)	(.028)	(.027)	(.017)
D2: Distance measure: pre-treatment frequency of achieving the target			
Quartile 1	Quartile 2	Quartile 3	Quartile 4 (>50%)
0.052	0.048	0.026	-0.009
(.022)	(.025)	(.030)	(.016)

Note: 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. Standard errors are clustered by shop. In Panel A, the samples are split into below and above the median age of the workforce excluding workers employed in a mini-job. 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 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.

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 ($\rho \gg \frac{1}{2}$). 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

$$\left| \Phi'' \left(a N^{\frac{1}{\rho}} e^* - y_0 \right) \right| \ll \Phi'(a N^{\frac{1}{\rho}} e^* - y_0),$$

$$\frac{de^*}{dN} = - \frac{a \cdot B N^{\frac{1-3\rho}{\rho}}}{\rho} * \frac{(1-2\rho)\Phi'\left(aN^{\frac{1}{\rho}}e^* - y_0\right) + N^{\frac{1}{\rho}}e \cdot a\Phi''\left(aN^{\frac{1}{\rho}}e^* - y_0\right)}{\frac{d^2\pi}{de_i^2}} < 0 \qquad \text{(A1)}$$

when $\rho > \frac{1}{2}$. For the total effort,

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

$$= e^* - \frac{a \cdot B N^{\frac{1-2\rho}{\rho}}}{\rho} * \frac{(1-2\rho)\Phi'\left(aN^{\frac{1}{\rho}}e^* - y_0\right) + N^{\frac{1}{\rho}}e \cdot a\Phi''\left(aN^{\frac{1}{\rho}}e^* - y_0\right)}{\frac{d^2\pi}{de_i^2}}, \qquad \text{(A2)}$$

whose sign is ambiguous. It can be shown that when output is linear in effort (no complementarities, $\rho = 1$), $\Phi(x) \approx x$, 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). Indeed, normalising quantities to suppress the inessential parameters a, b, B and y_0,

$$\pi(e_i, e_{-i}) = \frac{1}{N}\left(e_i + \sum_{j \neq i} e_j \right) - e_i^2 \qquad \text{(A3)}$$

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 $c(x) = x^k$,

$$\pi(e_i, e_{-i}) = \frac{1}{N}\left(e_i^\rho + \sum_{j \neq i} e_j^\rho \right)^{\frac{\gamma}{\rho}} - e_i^k, k > 1 \qquad \text{(A4)}$$

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

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

Prediction 5: Team total effort decreases with the share of non-incentivized members in the team θ. Let us temporarily ignore the probability of meeting the target as a function of effort. When θ goes up, the share in the total output received by each incentivized worker increases, which may lead to a higher individual effort by these workers. However, the total effort increase will be lower than individual because of incentivized workers being replaced by non-incentivized ones. In fact, under some plausible parameter values – most importantly, under $\rho \leq 1$ (effort complementarity) – the total output will go down.

To show this formally, we strip our model of unnecessary complications (such as non-linear transformation of effort into output) and use an approximation of the total effort (equation 2) with its second-order Taylor series expansion around the team average effort level (the method also applied in Iranzo et al., 2008):

$$E(e_1, \ldots, e_N) = \left(\sum_{i=1}^{N} e_i^\rho \right)^{\frac{1}{\rho}} \approx \left(\bar{e} + \frac{1}{2}(\rho - 1) * \frac{var(e)}{\bar{e}} \right) N^{1/\rho} \tag{A5}$$

where $e \geq 1$ is the optimal effort level by incentivized workers (the non-incentivized worker effort is normalized to 1), $\bar{e} = (1 - \theta)e + \theta$, is the average effort and $var(e) = (1 - \theta)\theta(e - 1)^2$ is the effort variance.

Each incentivized worker's share in output is

$$\frac{\bar{e} + \frac{1}{2}(\rho - 1) * \frac{var(e)}{\bar{e}}}{1 - \theta} N^{\frac{1-\rho}{\rho}} \tag{A6}$$

Differentiating (A6) with respect to θ at $\theta = 0$ gives

$$\left(\frac{(\rho - 1) * (e - 1)^2}{2e} + 1 \right) N^{\frac{1-\rho}{\rho}} \tag{A7}$$

That is, under the incentivized workers' effort e being not too different from that or non-incentivized and at ρ close to 1, a small increase in θ from the base level of 0 may actually result in a positive individual effort response driven by a larger share of output given to each incentivized worker.

Turning to the total effort $E(e_1, \ldots, e_N)$, differentiating it with respect to θ at $\theta = 0$ gives

$$\left(\frac{(\rho - 1) * (e - 1)^2}{2e} + 1 - e\right) N^{\frac{1-\rho}{\rho}} \qquad \text{(A8)}$$

Here the positive effect on total effort from the increase in individual shares in the total output (equation A7) is offset by the negative effect of replacing incentivized workers with less productive non-incentivized ones. The expression in (A8) implies that the effect of the non-incentivized workers' share on output is unambiguously negative when efforts are complementary ($\rho < 1$), and may still be negative even for some $\rho \geq 1$ if the difference between the incentivized and non-incentivized worker efforts, $e - 1$, is large. This negative effect is further exacerbated by the incentivized workers' anticipating a smaller chance of meeting the target - the factor we have so far ignored - and reducing their effort accordingly.

Prediction 6: The effort under the bonus will depend on the frequency of reaching the targets in the past, without the bonus. Less successful teams' effort response to the bonus will be stronger than that of more 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: the higher y_0, the lower is the probability of reaching it with the effort e_0. Consider first the interior solution case, when $e_0 < e^* < e_{max}$.

$$\left.\frac{de^*}{dy_0}\right|_{e=e_0} = -\frac{\frac{d^2\pi}{de_i dy_0}}{\frac{d^2\pi}{de_i^2}} = aN^{\frac{1-2\rho}{\rho}} B \frac{\frac{d\phi\left(aN^{\frac{1}{\rho}}e_0 - y_0\right)}{dy_0}}{\frac{d^2\pi}{de_i^2}} = -aN^{\frac{1-2\rho}{\rho}} B \frac{\phi'\left(aN^{\frac{1}{\rho}}e_0 - y_0\right)}{\frac{d^2\pi}{de_i^2}} \qquad \text{(A9)}$$

The derivative in (A9) is positive when the output, $aN^{\frac{1}{\rho}}e_0$, is at or below the target, y_0, since $\phi'(x)>0$ for $x < 0$. Thus, the less successful a team has been, the more effort it will put under a given bonus. However, the corner solution $e^* = e_0$ will be chosen by some very unsuccessful teams 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.