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

First version, preliminary and incomplete

For submission to the NBER workshop on Organizational Economics only

Guido Friebel, Matthias Heinz, Miriam Krüger, Nick Zubanov

Abstract: While there is ample evidence that incentive pay increases the performance of workers when individual performance can be measured, comparable evidence for teams is scarce. We fill this gap by reporting the results of a randomized experiment on team incentives with a retail chain of roughly 200 shops and 1200 employees. It is technologically impossible to measure individual performance, but the firm traditionally measures team, that is, shop performance along various dimensions. In 50% of the shops, we introduced a team bonus conditioned on targets fixed well before the team incentive was discussed. Treated shops increase their sales on average by three percent, and wages increase around 1.5 percent. Many of the shops reach sales levels beyond which the bonus is capped. Profits in the treated shops increase by a two digit number, and some workers increase their wages by ten percent. There is substantial heterogeneity in the treatment effect. Shops in cities compared to smaller communities feature much higher treatment effects, arguably because consumers react more intensively to increased sales activities and there is a broader range of products available in cities. Measures of job satisfaction seem to play no role.

Keywords: randomized experiment, randomized controlled trial (RCT), field experiment, team bonus, insider econometrics

JEL codes: D23, J33, M52

[1]All authors are at Goethe University Frankfurt, except for Heinz who is affiliated with the University of Cologne. Friebel is also affiliated with CEPR and IZA, and Zubanov with IZA. We are grateful for the support of Deutsche Forschungsgemeinschaft (DFG), and the comments of participants in seminars at Goethe University, Frankfurt, and the European Bank for Reconstruction and Development, London. We would also like to praise the team spirit of our partners in the retail chain, and of Artur Anschukov, Sandra Fakiner, Larissa Fuchs, Andre Groeger, Daniel Herbold, Malte Heisel, Robin Kraft, Stefan Pasch, Jutta Preussler, Elsa Schmoock, Patrick Schneider, Sonja Stamness, Sascha Wilhelm, who provided excellent research assistance.

1. Introduction

"How can members of a team be rewarded and induced to work efficiently?" This is the question that Alchian and Demsetz (1972) asked more than 40 years ago in an article that became one of the most influential contributions to the economic literature on organizations. Alchian and Demsetz suggested input monitoring, but an equally intuitive option would consist in providing team members with incentives conditioned on output. There is substantial evidence that incentives work very well, provided individual performance is measurable, cf. Lazear's (2000) influential study about workers in a windshield factory. The very nature of teamwork, however, blurs the performance of individuals into a common performance signal. Holmström (1982) shows that in order to overcome free riding in teams and reach first-best effort levels, all members would need to be penalized unless output reaches the efficient level. In reality, labor law, trade union and worker councils, and fixed wages make the use of such contracts impossible. Hence, it remains an empirical question whether or not team incentives can raise efficiency, and by how much.

Indeed, Bloom and van Reenen (2011, p. 1734) argue that there "appear(s) to be evidence that group incentive schemes also raise productivity, which is surprising given the free rider problem. Overall, in our opinion, however the evidence is weaker here than that for individual incentive pay." Our paper addresses this research gap on the effectiveness of team incentives through a randomized, controlled experiment in collaboration with a retail chain in Germany.

The retail chain consists of shops where employees bake and sell pre-fabricated bread and cakes, prepare and sell sandwiches, snacks, and hot and cold beverages. On average six employees work in each shop, most of them part-time workers with low levels of education or vocational training. Wages are low, at a level slightly above the currently debated, but not yet implemented nation-wide minimum wage. Individual performance measurement is impossible for technological reasons. There are a broad variety of tasks each person has to carry out, including handling the goods delivered, preparing food in the oven, taking care of the customers, handling the cash register. The time workers spend on each task varies much, as people work in overlapping shifts and are supposed to help each other. Providing individual incentives would necessarily lead to measurement and gaming problems and productivity losses (cf. Itoh, 1991, Auriol et al, 2002). However, for many years prior to our intervention, shop performance has been measured along a number of dimensions, such as sales, personnel costs, and qualitative indicators, all of which have traditionally been used to incentivize top,

middle and shop managers. Prior to the intervention, however, more than 1,000 sales agents in the shops never received any performance related pay.

In April 2014, in half of the almost 200 shops, we introduced a monthly team bonus. Shops were assigned to treatment and control groups through an optimal stratification procedure developed by Barrios (2014). The teams in treatment shops were informed through letters and posters about the incentive scheme, which entailed a step bonus function (Figure 1). Teams that reached or surpassed by up to one percent the sales target defined by top management at the beginning of the year (well before the decision in favor of the team incentive was made) would receive a 100 Euro bonus. For each additional one percent of sales beyond the target an extra bonus of 50 Euros was offered, with the total bonus capped at 300 Euro when the sales exceeded the target by more than 4 percent. Teams were initially informed that the bonus would be paid for a pilot period of three months, until June 2014 and that the team bonus would be shared among the full- and part time employees in the shop according to the hours they worked in the respective month.

We find a treatment effect of 3 percent of sales over the period between April and June 2014. Many of the shops reach sales levels beyond which the bonus is capped. Interestingly, and contrary to the free-riding argument, larger shops in the treatment group fare no worse in terms of sales than smaller ones. There is yet substantial heterogeneity in the treatment effect by shop location. Thus, shops in cities compared to smaller communities feature much higher treatment effects, arguably because consumers react more intensively to increased sales activities and there is a broader range of products available in cities. Measures of job satisfaction as generated by a firm-wide survey prior and unrelated to the team bonus seem to play no role. The firm made extra profits of around 50,000 Euros per month from the treated shops, and the wage payments to the sales agents increased by around 12,000 Euros per month. Following a successful implementation of the team bonus, the firm decided to roll out the scheme to all shops for the period from July to December 2014, signing the respective agreement with the worker council at the end of June.

We believe that retailing is a natural candidate to test for the effectiveness of team incentives. Many people in retail chains work in teams, that is, the individual efforts can only be mapped into some joint output signals, such as sales.[2] Demand volatility is high, so people must be prepared to switch tasks frequently, and even if sales people can specialize to some extent, they must be willing to help each other, all of which makes it complicated, if not impossible, to provide them with individual incentives. A study of team incentives study in

[2] http://job-descriptions.careerplanner.com/Retail-Salespersons.cfm

retailing, such as ours, will also make a general case because retailing is one of the largest industries in the world in terms of employment. *Wal-Mart* alone has more than 2 million employees. In Germany, the country in which our experiment took place, more than 4 million people work in retailing, that is, almost 10% of the country's active labor force. Most of these people work in larger groups or chains,[3] just as in our study. Other service industries, in particular fast food restaurants, are similar because, again, technology forbids individual incentives, people carry out many tasks, individual performance measurement is difficult, if not impossible, and people do not get individual incentives.[4]

Our paper contributes to the existing literature on incentives by providing clean causal evidence on the performance effects of team incentives, thus filling an important gap in empirical research. In our study, all shops have the same technology, team incentives are randomly assigned, people cannot self-select into teams, because hiring is centralized, and there are no moves between shops. This makes our study different from other papers in which the adoption of teamwork or team incentives may be endogenous. Boning et al (2007), Hamilton et al (2003), Bandiera et al (2009) and Bandiera et al (2013) all find supportive evidence that team incentives raise efficiency, but the first study shows that the decision in favor of teamwork and its effects depend on technology, while the others observe sorting of workers of similar productivities into teams. Ichniowski et al (1997) have shown that there exist strong complementarities between different HR systems, which makes it important to measure the joint changes of all HR measures to estimate causal effects. We have taken great care to only vary one aspect of the HR system between treatment and control groups, the team bonus, while all other components were kept fixed.

2. Background

In 2013, we were contacted by the general manager of a bakery chain who sought advice on how to cope with the challenges of a rapidly changing market. Since the 1980s, bakery chains, some of them owning hundreds of shops and with sales numbers of up to a billion Euro, had successfully built their business model on attractive locations (including 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, discounters *Aldi* and *Lidl* had begun to bake and sell fresh bread, rolls and related products in

[3] http://de.statista.com/statistik/daten/studie/261517/umfrage/beschaeftigte-im-deutschen-einzelhandel-nach-unternehmensgroessen/

[4] Other type of restaurants and cafés a very different type of business. Here, workers are much more specialized and they receive individual incentive pay in the form of tips.

their dense network of shops, with large success. Their bread is widely believed to be of similar quality as the one of the chains, but sold at much lower prices, hence forcing the incumbent chains to rethink their business model. As a consequence, many of the chains were differentiating their product portfolio moving into the market for snacks and sandwiches and beverages, traditionally covered by cafés and fast food chains.

The manager told us that for this strategic move to be successful, the behavior of the personnel should be changed, and that they needed to be more actively involved in sales activities. The English saying "something sells like hotcakes" has a German equivalent "something sells like warm rolls" or "something sells like sliced bread", and, indeed, many of our partners in the firm ("the bakery") complained that sales agents took a rather passive stance in their interactions with customers. Moreover, turnover in the bakery was high, making training activities a questionable investment, also because personnel consisted mainly of very low-skilled workers without comprehensive vocational training (only one fourth of the staff have finished an apprenticeship). The bakery has experimented before with hiring more qualified employees, but had to realize that there was little if any increase in the quality, but suffered from a substantial increase in the wage bill. So, the challenge was to find a lever through which the existing staff could be motivated to spend more effort on their sales activities.

Before working on finding the right instrument, we agreed to help the bakery under the condition that we would have full access to all relevant data, and could design any possible intervention through a randomized controlled experiment. In particular, we explained that our randomization would be more successful, if we had access to historic sales data. We received sales data over more than two years, allowing us a very precise stratified assignment, which is explained in more detail in the next section. In exchange we offered our advice free of charge.

When our partners presented the existing compensation structure to us, we realized that there was a very detailed system of key performance indicators (KPI) according to which managers were evaluated. These KPIs include sales, personnel costs, specific strategic goals such as enhancing the share of a certain product in total sales etc. Most importantly, at the end of each year, sales targets are determined for the next year, and these targets are never adjusted during the year. For the year 2014, management had based targets on 2013 sales, corrected by an expected decrease of 2% on average. District managers, who are responsible for 10 to 15 shops, are incentivized through a number of targets along these dimensions, and the 200 shop managers receive bonuses when they reach certain sales and personnel cost

targets and certain grades in regularly carried out mystery shopper visits (the topic of another paper of ours). However, sales agents, representing the bulk of staff, only receive fixed wages which for most workers are regulated by collective agreements.

When at the end of February 2013 we proposed introducing a team bonus conditional on reaching or passing the existing sales target, our partners first reacted in a less than enthusiastic way. A member of the management team put it bluntly: "We have never seriously thought about this." Other members of the management team were afraid that payments could turn out to be a burden on the firm. We provided our partners with some simulations showing that the expected bonus payments would likely be lower than 20,000 Euro per month in case half of the shops were treated. Further, we suggested a step function with a cap at 4% sales above the target providing a bonus of 300 Euro for a team. While this convinced top managers to try a "pilot" study with half of the shops assigned to the team bonus, district managers complained because raising wage costs would mean that they could not reach the targets of keeping personnel costs low. The general manager reacted by suggesting that the extra bonus payments would be paid from a different budget and would not affect the personnel costs relevant for district managers' performance. District managers were quick to realize that in such a setting they were likely to benefit as well, if the team bonus resulted in an increase of sales of the shops under their supervision.

The next concern raised was about the possibility of envy between treatment and control shops. We discussed this issue with the worker council who suggested that we should explain that the intervention was "just a pilot" and that everybody would have the same 50-50 chance of taking part in it. This, the worker council argued, would be acceptable for the non-treated shops in case they would learn about the bonus scheme in other shops. In the next section, we explain in more detail how we communicated the scheme and how we minimized the risk of contamination between the treated and non-treated shops.

3. The experiment

In March 2014, prior to the introduction of the team bonus in April 2014 and before its announcement to the treatment group of shops, we carried out an employee survey in order to measure employee satisfaction with the job context, general situation, and organizational commitment, following some influential work in industrial psychology (Allen and Meyer, 1990). We also carried out a second wave of an employee survey during the month of June. The latter survey entailed the same questions plus additional items in relation to the team bonus. Our partners and we thought the survey would be useful for three reasons. First, it

would allow an additional check whether treatment and control groups were balanced samples; second, we hypothesized that the treatment effect may depend on employees' satisfaction levels; third, it would be likely that because of the team bonus, employees' satisfaction levels could increase. The survey was distributed through the district managers and collected by our research assistants.

In preparation of team bonus introduction, we designed information leaflets to be posted in the back offices of the treatment shops, and letters that were 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 2) because we wanted to avoid people to speculate about the link between the employee survey that was visibly carried out by the team of Goethe University, and the team bonus, a policy of he management of the bakery. In particular, with a view to the second wave of the survey, we decided not to reveal our engagement.

We trained district managers for explaining the team bonus to shop managers and instructed them how to react to questions of employees in control group shops about the bonus. In that case, district managers would explain that "this is a pilot. Everybody had the same chance to be drawn with a 50-50 chance. The work council agreed to this procedure." We were afraid that there could be contamination between the treatment and control group. Envy or frustration of control group shops could lead them to reduce their efforts, which could then be picked up erroneously as a positive treatment effect. To monitor this risk, we regularly called the district managers and asked them whether employees in the control group had asked them questions about the team bonus. We actually only heard about three cases, all of which were satisfied with their district manager's answer. However, we also regularly checked the Facebook page of the bakery on which customers and employees alike tend to discuss issues such as product quality but also (sometimes to the dissatisfaction of management) internal issues such as stress at the workplace, quality issues of products, or problems of leadership and organizational culture. We could not find a single entry on the team bonus. However, we also built in some questions into the second wave of the employee survey to check for potential channels of contamination, in particular, we asked to what extent people tend to interact with colleagues in the same and in other shops, and we found that contacts to employees in other shops are very rare.

FIGURE 1 ABOUT HERE

Figure 1 depicts the bonus offered to the treatment shops. Shops that reach the sales target for the month receive a bonus of 100 Euro to be shared between the part-time and full-

time employees in the shop. The bonus increases by 50 Euro for each percent point above the target and is capped at 300 Euro per month at 4 percent points above the target. Hence, the team in a shop can make extra earnings of up to 900 Euro in the treatment quarter April to June. We provided the employees with examples of what sales increases would mean in terms of additional goods to be sold per day (for instance a one-percent increase above the sales target would be tantamount to selling per day ten additional rolls, two loafs of bread, some sandwiches and some cups of coffee for mid-size shop). The fact that a number of shops failed to reach the target by small amounts is an indicator that there was no manipulation and that at least in the beginning of the treatment phase, employees found it hard to estimate their likelihood to reach the various targets, although district managers regularly communicate to all shops, treatment and control shops alike, their current sales figures.

What rests to be explained is the randomization procedure. We follow Barrios (2014) who shows that randomizing pairwise by using the predicted outcome variable, in our case sales, minimizes the variance of the differences in the treatment and control outcomes post treatment. We use historic observations to run a regression of log sales on an array of observables such as location, shop size, workforce characteristics, etc, to obtain predicted sales, and then rank the shops according to the predicted sales. We randomize within the pair of shops with adjacent ranks (1-2, 3-4, and so on).

FIGURE 2 ABOUT HERE

The resulting assignment of 96 shops to the control and 97 shops to the treatment group is shown on the map (Figure 2) and the descriptive statistics by group are reported in Table 1. The region of relevance is roughly 100 km from West to East and 60 km from North to South, a mini economy of ca. 2 million inhabitants. As a result of our randomization procedure, we have the treatment and control samples balanced in terms of the average pre-treatment sales as well as other potentially relevant characteristics. In fact, none of the averages reported in Table 1 differ significantly between the groups. An average shop sells ca. 28,000 Euros worth of goods, employing 6 staff most of whom are female in their late 30s working part-time. The sales are quite variable: while location and size differences explain the lion's share of variation in sales between shops, there is also a considerable variation within shops, much of which we were told is due to seasonal demand, temporary closures for renovation, and market dynamics, such as the entry and exit of competitors.

TABLE 1 ABOUT HERE

4. Results

To inspect the effect of the team bonus, we first plot the year on year sales growth for the months April, May, June on the log of sales in the year before, in Figure 2. Treatment shops have higher sales growth then control shops as indicated by the average lines. The Kernel distribution of the three months are plotted in Figure 3, the treatment distribution is almost everywhere right from the control group.

FIGURES 2 AND 3 ABOUT HERE

To identify the treatment effect in a more precise way, we start with running the basic difference-in-difference estimator with the following regression equation:

$$(1)\ \ln(sales_{it}) = \beta_0 + \beta_1 \cdot treatment_i + \beta_2 \cdot after_t + \beta_3 \cdot treatment_i \cdot after_t + error_{it}$$

where $\ln(sales_{it})$ is the log sales in shop i and month t, the *treatment* dummy takes the values 1 or 0 for the treatment and control group shops, respectively, the *after* dummy is 1 or 0 for the periods after and before the start of the treatment, and $error_{it}$ is the regression error term. Coefficient β_3 is the difference-in-difference estimate of the treatment effect. The results in columns 1 and 2 of Table 2, which reports our baseline regression results, are based on equation (1).

TABLE 2 ABOUT HERE

There are two potential estimation issues with the difference-in-difference estimators. First, there may be serial correlation in the residuals, which, when ignored, leads to underestimated coefficient standard errors (Bertrand, Duflo and Mullainathan, 2004); second, despite randomization, in finite samples, there may be correlation between the treatment status and baseline outcomes, causing a "regression to the mean" problem (Stigler, 1997).

To address the first issue, we routinely cluster regression standard errors by shop. In addition, as another solution discussed in Bertrand, Duflo and Mullainathan (2004), we run a specification in which we collapse all observations into the pre- and post-treatment periods:

$$(2)\ \ln(sales_{i\tau}) = \beta_0 + \beta_1 \cdot treatment_i + \beta_2 \cdot after_\tau + \beta_3 \cdot treatment_i \cdot after_\tau + error_{i\tau}$$

where the dependent variable, $\ln(sales_{i\tau})$, is the average log sales pre and post treatment, and index τ is 0 pre and 1 post treatment. Column 3 in Table 2 reports the treatment effect estimated with equation (2). Bootstrapping, another recommended solution, produces standard errors of similar magnitude.

To control for regression to the mean in we augment eq. (2) with past sales:

$$(3)\ \ln(sales_{i1}) = \beta_0 + \gamma \cdot \ln(sales_{i0}) + \beta_1 \cdot treatment_i + \beta_2 \cdot after_\tau + \beta_3 \cdot treatment_i \cdot after_\tau + error_{i\tau}$$

The estimates from equation (3) are reported in column 4 of Table 2 column 4.

In the next series of specifications, we calculate the treatment effect by comparing post-treatment sales growth relative to a chosen baseline *b* in the treatment and control groups while continuing to control for regression to the mean with the baseline sales. The general form of the regression equation we run is:

$$(4)\ \ln(sales_{it}) - \ln(sales_{ib}) = \beta_0 + \gamma \cdot \ln(sales_{ib}) + \beta_3 \cdot treatment_i + error_{i\tau}$$

In principle, doing so is very close to running equation (3), but some extra flexibility in regression specification is achieved by varying the baseline. We use different specifications for the baseline sales, $\ln(sales_{ib})$: the average across all months before the start of the treatment (column 5 in Table 2), the same months in 2013 (column 6), the average sales in April-June 2013 (column 7), and the average sales in January-March 2014 (column 8).

Whatever specification we use, we obtain statistically significant effects of the team bonus on team (shop) performance. The point estimates in most of our specifications are around 3%, varying from 3.5% to 1.8%. Calculating the treatment effect in each month by running (3) with the average pre-treatment sales used as the baseline, we find it to be 2.7% in April 2014, 3.6% in May, and 3.2% in June, a steady effect without noticeable abatement.

We should expect treatment effects to be heterogeneous along at least two dimensions: shop size and location. First, one could expect shops with fewer employees to feature stronger reactions to the team bonus, compared to shops with more employees, as predicted by the free-riding effect. Surprisingly, the regressions we have run do not provide support for this prediction. The interaction of treatment with quartile dummies of shop size by employment is insignificant, suggesting no systematic differences in the treatment effect with shop size.

TABLE 3 ABOUT HERE

A second dimension, however, turns out to be more interesting. Differentiating treatment effects by location (large towns, smaller towns, villages) shows a very clear pattern, as documented in Table 3. Only in larger towns do we find a statistically significant effect of 6.5 percent, which is remarkably stable over the three months considered. The point estimates of smaller towns are positive, the ones of villages negative, but both are statistically insignificant.

One hypothesis in line with this result is that in larger towns, customers' sales decision may be more sensitive to employees sales efforts, for two reasons. First, in smaller town and villages, there are less business people to whom one can upsell snacks and beverages, in particular over lunchtime, while in smaller municipalities, people may have more habitual shopping patterns. A second explanation may be on the supply side: shops in towns and cities

may have a broader range of products such that there are more opportunities to upsell attractive products.

TABLE 4 ABOUT HERE

While we have not yet been able to control for the second explanation, we made an attempt to control for the first one. Our research assistants visited 140 of the shops (owing to time constraints we could not visit all shops). They were instructed not to reveal themselves as being a member of the research team and had a simple task, buy the „bread of the month" or the closest substitute to it, and after leaving the shope take note of (i) the friendliness of staff, (ii) whether or not staff had asked the standard question „Would you like anything else?" or a similar question. We found a striking difference, as documented by Table 4. There is no difference in the frequency of asking such a question in smaller towns and villages, but for larger towns, in the treatment group, sales staff are 20 percent more likely to ask the question. It hence seems that the treatment is inducing more proactive sales behavior when sales staff believe it to be useful. However, while regressions in Table 5 show positive, albeit insignificat, effect of asking the „Anything else?" question on sales, the treatment effect is not mediated by proactive sales behavior. Indeed, when we include the „Anything else?" dummy in the sales equation together with the treatment dummy (Table 6), its estimate is small and insignificant, whereas the treatment effect remains of the same magnitude and significance as before.

TABLE 5 ABOUT HERE

TABLE 6 ABOUT HERE

Many more analyses are possible, in particular, we have not yet systematically used the results of the pre- and post treatment survey, and in terms of left-hand side variables, we have not yet coded absence and turnover, factors of great interest for the firm. What we have done, though, is computing the gains from introducing the team bonus for the firm and the employees based on the April 2014 data (the relevant data for the later months are still being processed). The treatment effect of 2.7% in April implies extra sales of 770 euro on average per shop in April times 96 shops in the treatment group = 73,920 Euro per month. The actual total bonus payment for April was 11,500 Euro on top of the usual personnel costs. Many shops actually reached the top bonus of 300 Euro. Given the historic share of value added in sales of 56%, the extra value added is 40,500 Euro per month. There are indirect incentive costs for incentive schemes (for shop managers, district managers, head of distribution) that we estimate at 5,000 Euro. Assuming no further increase in personnel costs, the net profit gain is 40,500 - 11,500 – 5,000 = 24,000 Euro per month when half of the shops are treated. For a

full cost perspective, our time cost of research should be billed at a reasonable rate (though we were not paid). We estimate around 50 person days of senior researchers at a rate of 1,000 Euro = 50,000 Euro. Also we estimate 50 research assistant days at the going rate of 110 Euro = 5,500 Euro. Material and travel cost were around 12,000 Euro. So, at implementation cost of approximately 67,500 Euro, and a roll out to all shops, the project break-even would be reached after 1.5 months.

In a meeting in June 2014, the management team decided to roll out the team bonus to all shops. A collective agreement was written with the work council, according to which the team bonus would be granted until 31st of December 2014.

5. Concluding discussion

Teams are a ubiquitous feature of modern production, and so are monetary incentives. While the knowledge about the effectiveness of individual incentives is both broad and deep, much less is known about team incentives. Problems of endogeneity and complementarities and self-selection into teams put the bar for causally interpretable evidence about the effectiveness of team incentives very high. We have designed a fairly large randomized controlled experiment with 193 shops of a bakery chain in Germany. Power calculations on the basis of 27 months of observations pre treatment and 3 months post treatment informed us that we would need 70 shops in each group to detect a 3 percent treatment effect at a significance level of 5 percent with the probability 0.9. Our estimated treatment effect is indeed around 3%, and is highly significant. There is also substantial heterogeneity, with a treatment effect of up to 7 percent in larger towns, and an insignificant effect in smaller towns and villages.

We carried out first inquiries into the channels of these effects, but found none yet. While there is some support that the team bonus made employees more likely to become more proactive in the interaction with customers, proactive sales behavior does not mediate the treatment effect. Other channels can, in principle, be examined. For instance, provided we will receive footing data, we could investigate whether employees shift their activities towards customers with higher propensity to buy. There are also promising left- and right-hand variables we have not yet fully investigated, for instance, personnel turnover, job satisfaction data before and during the experiment.

The decision of the firm to roll out the team bonus to all 193 shops is certainly part of the proof of concept, and we hope to gain additional insights from the data of the period between July and December 2014. The full roll out of team incentives brings about methodological complications, as we have no natural control group anymore. However, it can

be checked to what extent the course taken by our treatment group is similar to the one taken by the control group, once the team bonus is introduced.

The fact that the number of employees in a shop play little if any role in the performance is not in line with the predictions by theory, which says that a given bonus should have a weaker effect once team size increases. However, the monetary bonus we provided may have been a trigger for a number of other processes, in particular, an activation of peer pressure, which as shown theoretically by Kandel and Lazear (1992), experimentally by Falk and Ichino (2006) and in the field by Mas and Moretti (2009) can induce substantial efficiency effects. We will investigate to what extent the social connections at the workplace, as documented by our employee survey, matter for the effect of the team bonus.

References

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American Economic Review*, 62(5), 777-795.

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

Auriol, E, Friebel, G. & Pechlivanos, L. (2002). Career concerns in teams. *Journal of Labor Economics*, 20(2), 289-307.

Bandiera, O., Barankay, I & Rasul, I. (2013). Team Incentives: Evidence from a Field Experiment. *Journal of the European Economic Association*, 11(5), 1079-1114.

Bandiera, O., Barankay, I & Rasul, I. (2009). Social Connections and Incentives in the Workplace: Evidence from Personnel Data. *Econometrica*, 77(4), 1047-1094.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Difference-In-Difference Estimates. Quarterly Journal of Economics, 119(1), 249-275

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

Bloom, N. & Van Reenen, J. (2011). Human Resource Management and Productivity. *Handbook of Labor Economics*, 4B(19), 1697-1767.

Boning, B., Ichniowski, C. & Shaw, K. (2007). Opportunity Counts: Teams and the Effectiveness of Production Incentives. *Journal of Labor Economic*, 25(4), 613-650.

European Foundation for the Improvement of Working and Living Conditions (2007), Teamwork and high performance work organisation.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1), 39-57.

Hamilton, B., Nickerson, J. & Owan, H. (2003). Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. *Journal of Political Economy*, 111(3), 465-497.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 13(2), 324-340.

Ichniowski, C., Shaw, K. & Prennushi, G. (1997). The Effects of Human Resource Management Practices on Productivity. *The American Economic Review*, 87(3), 291-313.

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

Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy,* 100(4), 801-817.

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

Mas, A., & Moretti, E. (2009). Peers at Work. *American Economic Review*, 99(1), 112-145.

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

Appendix I: Tables and Figures

Figure 1: The Team Bonus

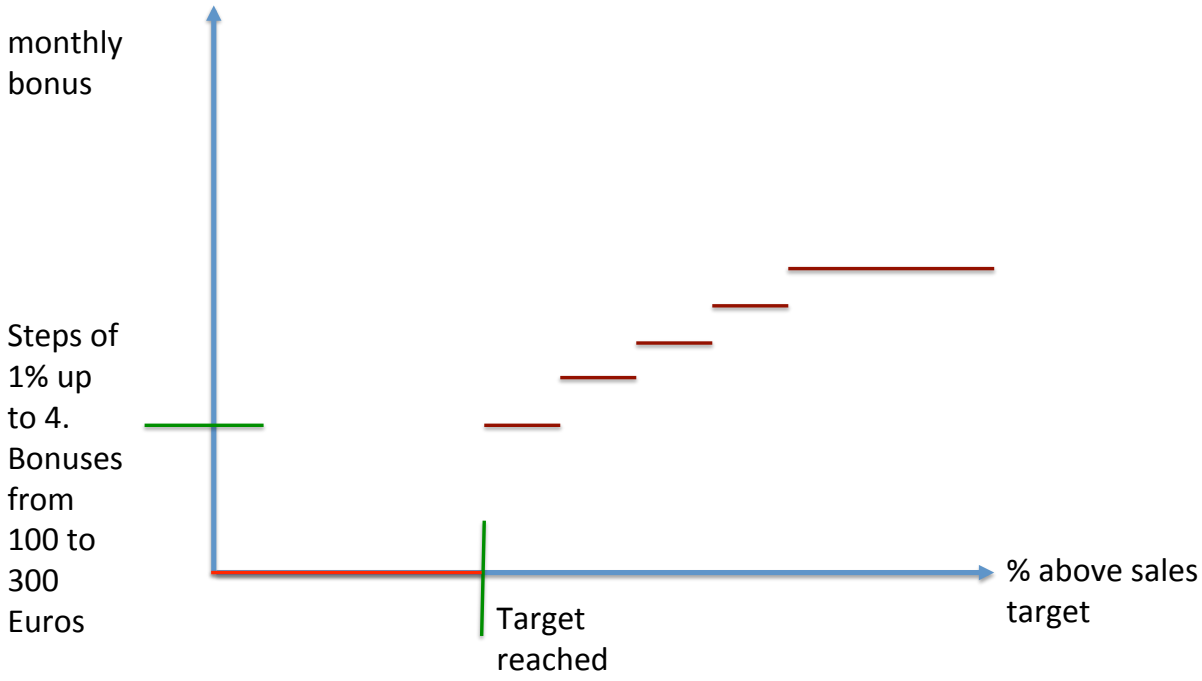


Figure 2: A map of shops by treatment (red) and control group (blue)

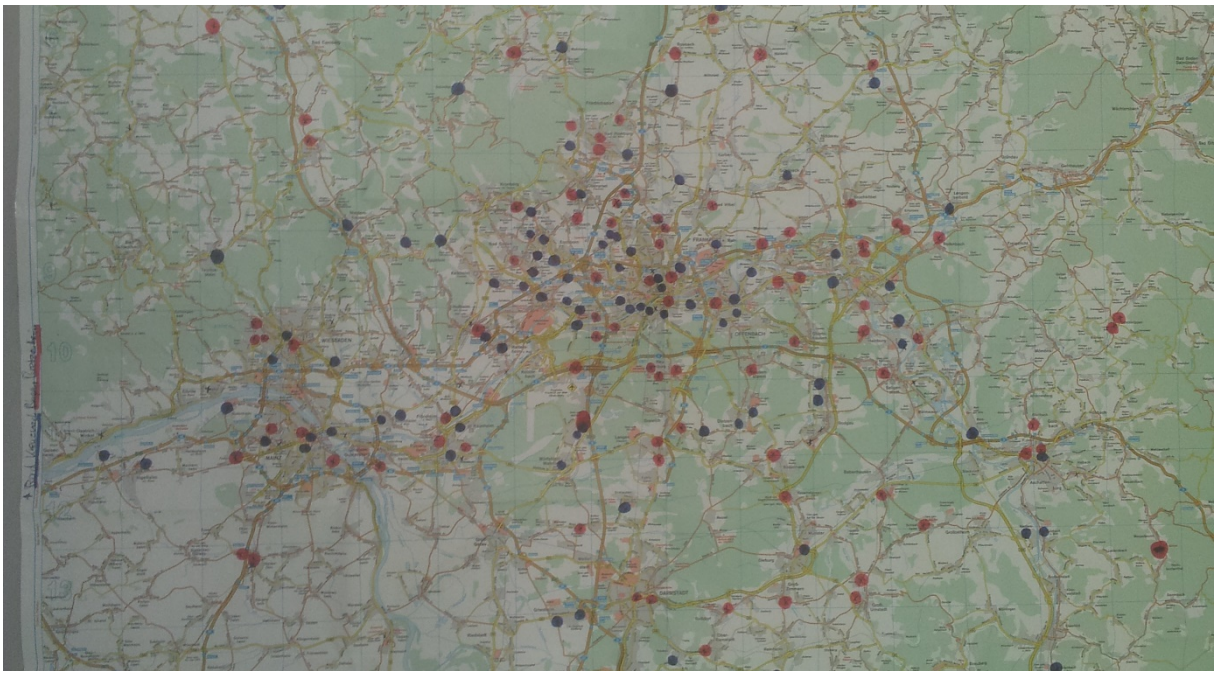


Figure 3: Scatter plot, year on year sales growth on log sales, April, May and June 2013

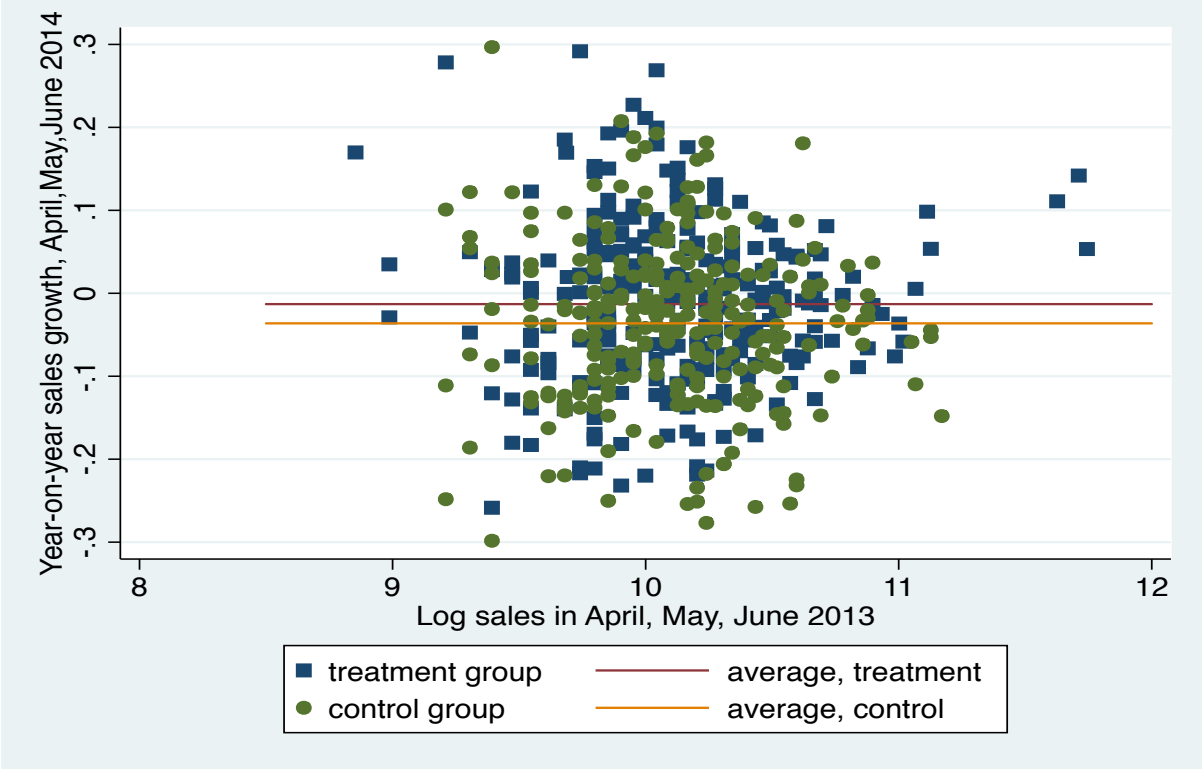


Figure 4: Kernel distribution sales growth treatment versus control group

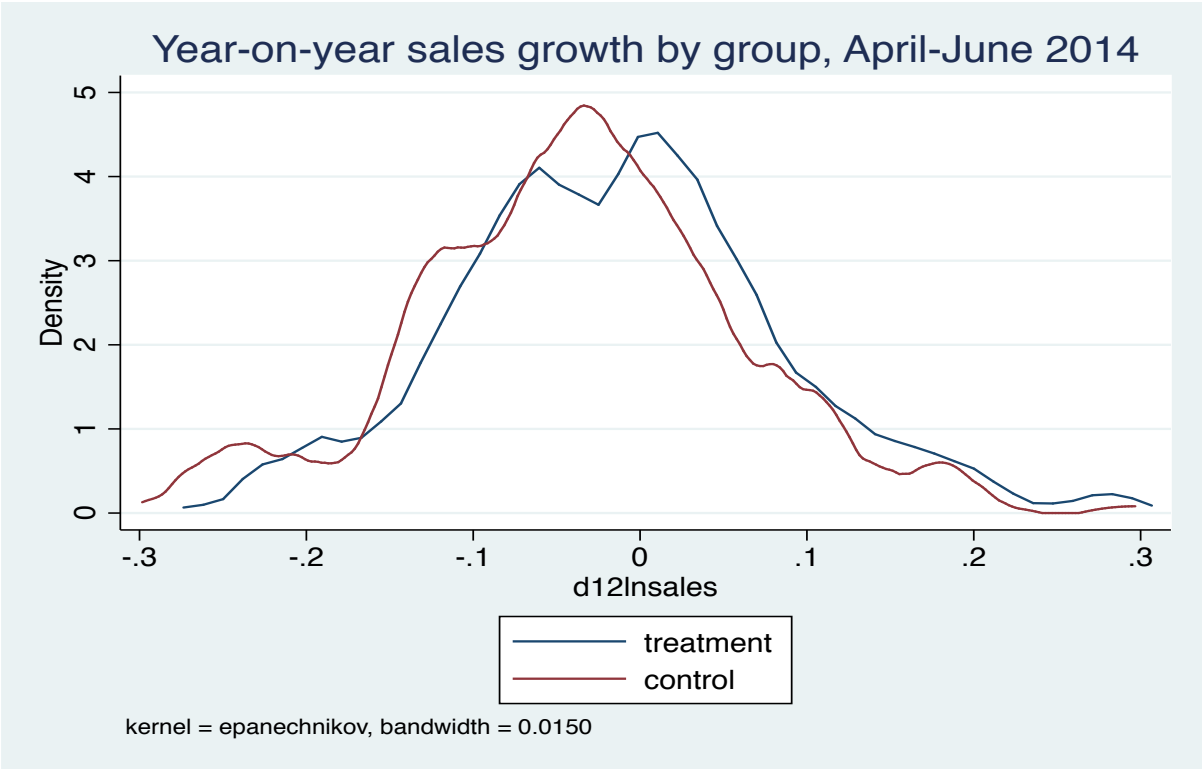


Table 1: Characteristics of the control and treatment shops before the treatment

	Panel A: Sales before implementation of the treatment		
	Control group (n = 96)	Treatment group (n = 97)	T-test (p-value)
Mean monthly sales (SD)	27453 (11481)	28194 (14542)	0.700
Mean monthly sales (in logs, SD)	10.143 (0.391)	10.154 (0.415)	0.846
	Panel B: Location of shop		
Big town	37.50%	34.00%	
Medium/small town	26.04%	29.89%	
Village	36.45%	36.70%	
	Panel C: Characteristics of shop managers		
Number of shop managers	96	97	
Age	42.15 (10.44)	42.98 (10.70)	0,586
Female	98,90%	96,90%	
Full time	91,70%	92,70%	
	Panel D: Characteristics of sales agents		
Number of sales agents	545	561	
Age	38.56 (13.81)	37.75 (14.29)	0.337
Female	96.15%	94.83%	
Temporary emplyoment (% of all sales agents)	42.39%	44.03%	
Full-time employment (% of all sales agents)	10.44%	8.72%	
Part-time employment (% of all sales agents)	53.59%	54.71%	
Marginal employment (% of all sales agents)	32.67%	34.58%	
Unskilled worker (% of all sales agents)	77.61%	74.33%	
	Panel E: Employee attitudes, based on questionnaire data		
Mean commitment score (SD)	4.500 (1.548)	4.424 (1.691)	0.523
Mean work satisfaction score (SD)	4.445 (1.514)	4.334 (1.567)	0,422
Mean overall satisfaction score (SD)	4.980 (1.623)	4.898 (1.701)	0,548

Panel A reports monthly sales between January 2012 and March 2014, the 27 months before implementation of the team bonus. Standard deviations are in parentheses. Column 3 provides the p-values of the two-sided t-test of equality of the means. Panel B shows the shares of shops located in big towns (90,000 or more inhabitants), medium/small towns (between 5,000 and 60,000 inhabitants) and villages (less than 5,000 inhabitants). Panel C and D are based on the personnel records from the firm as of April 1st 2014, omitting apprentices and interns (18 in the control and 11 in the treatment group). Panel E lists descriptive statistics on the commitment, work satisfaction and overall satisfaction indices following Allen and Meyer (1990) from the employee survey prior to the treatment. In total, 563 employees in the control, and 580 employees in the treatment group participated in the survey (response rate: 79.5%).

Table 2: Treatment effect estimates, April-June 2014

Specifications	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment effect	0.032	0.034	0.034	0.035	0.032	0.023	0.018	0.026
	(0.014)	(0.012)	(0.013)	(0.012)	(0.012)	(0.011)	(0.008)	(0.013)
Observations	5260	5188	386	182	549	543	543	536

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 72 outliers, defined as year-on-year changes in sales of more than 30%. 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. Specifications 3 and 4 are the same, 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 a control. In specification 5, we regress the difference between the log monthly sales in April, May and June 2014 (the treatment period) and the average sales in the respective shop in the months before the treatment on the treatment group dummy. In specification 6, we do the same with the difference between the log monthly sales in the treatment period and the respective months in 2013 as the dependent variable. Specification 7 is the same except that we use the difference in log monthly sales in the treatment period and the average sales in April-June 2013 as the dependent variable. We do the same in specification 8, except using the average sales in January-Mach 2014 as the baseline. Note that we have had to omit some observations because some shops were closed, either temporarily (renovation) or permanently. Standard errors are clustered by shop. Bootstrapped standard errors (available on request) are similar in magnitude.

Table 3: Treatment effect by shop location

Specifications	April 2014	May 2014	June	Overall
Big towns	0.065	0.073	0.051	0.064
	(0.023)	(0.023)	(0.025)	(0.021)
Small-/medium sized towns	0.001	0.035	0.004	0.015
	(0.023)	(0.025)	(0.028)	(0.019)
Villages	-0.012	0.000	-0.013	-0.010
	(0.017)	(0.022)	(0.018)	(0.015)

The specification that produces the estimates reported in this table is the same as specification 6 in Table 2: year-on-year changes in April, May and June 2014 regressed on the treatment dummy and the respective past sales. The cells in the table give estimated treatment effect in a given month and location. For example, 0.065 is the treatment effect in April 2014 in shops located in big towns. Standard errors are clustered by shop.

Table 4: Percentage of employees asking "Anything else?" or a similar question, by shop location

	Control shops	Treatment shops
Overall	0.718	0.795
Big town	0.618	0.821
Medium/small town, villages	0.783	0.778

The table provides the percentage of employees who asked the question: "Do you want to have anything more" or a similar question during our mystery shopping tour in May 2014.

Table 5: Proactive sales behavior and monthly sales, April-June 2014

Specifications	(1)	(2)	(3)
Question asked	0,068	0.069	0.068
	(0.088)	(0.052)	(0.051)
Shop size (in logs)		0.952	0.953
		(0.076)	(0.081)
Location controls	No	No	Yes
Observations	435	432	432

The dependent variable is the log monthly sales in April-June 2014. Question asked is a dummy set to one if the shop employees asked the student assistants on the mystery shopping tour whether they would like anything else. Standard errors are clustered by shop.

Table 6: Treatment effect controlling for proactive sales behavior

Specifications	Overall	Overall	Location: Big town	Location: Others
Treatment effect	0.031	0.029	0.065	0.005
	(0.012)	(0.012)	(0.023)	(0.013)
Question asked		0.025	0.037	0.007
		(0.016)	(0.029)	(0.015)
Observations	426	426	169	257

The specification producing the results in this table is the same as in Table 3. The first column shows the treatment effect on the mystery shopping subsample (148 shops) without controlling for whether the shop assistant asked "Anything else?" or a similar question. The "Anything else?" dummy us added in the second column. The third and fourth column list the regression results from the same specification as in the previous column run on the shops located in big towns and elsewhere. Standard errors are clustered by shop.

Appendix II

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.