American Economic Review 2017, 107(8): 2168–2203
https://doi.org/10.1257/aer.20160788

Team Incentives and Performance: Evidence from a Retail Chain†

By Guido Friebel, Matthias Heinz, Miriam Krueger, and Nikolay Zubanov*

In a field experiment with a retail chain (1,300 employees, 193 shops), randomly selected sales teams received a bonus. The bonus increases both sales and number of customers dealt with by 3 percent. Each dollar spent on the bonus generates $3.80 in sales, and $2.10 in profit. Wages increase by 2.2 percent while inequality rises only moderately. The analysis suggests effort complementarities to be important, and the effectiveness of peer pressure in overcoming free-riding to be limited. After rolling out the bonus scheme, the performance of the treatment and control shops converges, suggesting long-term stability of the treatment effect. (*JEL* D22, J31, J33, L25, L81, M53, M54)

"How can members of a team be rewarded and induced to work efficiently?" This classic question, posed by Alchian and Demsetz (1972, p. 779) in their influential contribution to the economic analysis of organizations, lies at the heart of this paper.

* Friebel: Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany, CEPR, and IZA (email: gfriebel@wiwi.uni-frankfurt.de); Heinz: University of Cologne, Albertus-Magnus-Platz, 50923 Koeln, Germany, and CEPR (email: heinz@wiso.uni-koeln.de); Krueger: Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt, Germany (email: miriam.krueger@bundesbank.de); Zubanov: University of Konstanz, Universitaetstrasse 10, 78464 Konstanz, Germany, and IZA (email: nick.zubanov@uni-konstanz.de). We are grateful for the support of Deutsche Forschungsgemeinschaft (DFG). We would like to thank for their comments: Oriana Bandiera, Iwan Barankay, Stefan Bender, Nick Bloom, Viv Davies, Stefano DellaVigna, Thomas Dohmen, Florian Englmaier, Niels Kemper, Michael Kosfeld, Johan Lagerloef, John List, Jan Luksic, Hideo Owan, Allison Raith, Michael Raith, Imran Rasul, Werner Reinartz, Devesh Rustagi, Kathryn Shaw, Raffaela Sadun, Heiner Schumacher, Bruce Shearer, Orie Shelef, Dirk Sliwka, Matthias Sutter, Ferdinand von Siemens, Etienne Wasmer, and participants in seminars at Adelaide, NHH Bergen, WZB/TU Berlin, Birmingham, Bonn, QUT Brisbane, Cologne, Columbia, Copenhagen University, CUNEF Madrid, Dunedin, University of East Anglia, the EBRD, Innsbruck, King's College, Konstanz, Luxembourg, Maastricht, Sciences Po Paris, Queen's, Rotterdam, UNSW Sydney, Trier, and conferences organized by the University of Arhus, the LMU Munich, the annual GEABA meeting in Regensburg, the COPE conference in Vienna, the NBER Organizational Economics Working Group meeting 2014 in Stanford, the 2015 meeting of the Academy of Management in Vancouver, the 2015 CEPR IMO conference at INSEAD, the 2015 RCT conference in Paris, and the MWO conference 2015 in Madrid. We would like to praise the team spirit of the partners in the study firm, and of Artur Anschukov, Sidney Block, Sandra Fakiner, Larissa Fuchs, André Groeger, Daniel Herbold, Malte Heisel, Robin Kraft, Stefan Pasch, Jutta Preussler, Elsa Schmoock, Patrick Schneider, Sonja Stamness, Carolin Wegner, Sascha Wilhelm, and Sandra Wuest, who provided excellent research assistance. We are grateful to *ImmobilienScout24* who provided us with data within the framework of their transparency initiative on the real estate market in Germany. We did not obtain an IRB approval for this project because Goethe University did not have an IRB at the time the experiment was carried out; at that time, the entire research team was employed at Goethe University. However, we would like to stress that the firm's work council approved and supported the project. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. Any opinions expressed in this paper represent the authors' personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank or its staff.

† Go to https://doi.org/10.1257/aer.20160788 to visit the article page for additional materials and author disclosure statement(s).

Alchian and Demsetz (1972) argued in favor of input monitoring of employee performance by a manager-owner. The natural alternative to this would be incentives conditioned on joint output. However, teamwork blurs the performance of individuals into a common signal, which can result in free-riding by individuals that weakens the effectiveness of team incentives (Holmström 1982). In the presence of managerial instruments other than the team incentives, team incentives may fail entirely in providing additional motivation, and hence lead to no efficiency gain at all.

To establish whether team incentives can substantially increase performance presents a significant and important empirical challenge. Teamwork is a ubiquitous feature of the modern economy (Deloitte 2016), and team incentives are gaining importance in the global economy, which is undergoing a shift from manufacturing toward services. Yet, unlike for the case of individual incentives (Lazear 2000; Shearer 2004; Bandiera, Barankay, and Rasul 2009), the jury on the effectiveness of team incentives is still out (Bloom and Van Reenen 2011). This leaves a considerable gap in the current research on this subject.

At the core of this gap lie two identification problems that need to be solved in order to generate causally interpretable evidence on the effectiveness of team incentives. The first is that individuals may self-select into treatment, a fundamental issue for the identification of HR practices' performance effects in general, and team incentives in particular (Prendergast 1999; Hamilton, Nickerson, and Owan 2003; Bandiera, Barankay, and Rasul 2013). The second, also discussed by Prendergast (1999), is that, across firms, technology and profitability differ, and that these differences are relevant for decisions in favor of team organization and compensation (Boning, Ichniowski, and Shaw 2007).

To solve these issues, we design a field experiment which combines randomization and realism (Harrison and List 2004; List and Rasul 2011). We introduce a team bonus[1] for a randomly selected half of the shops of a bakery retail chain with 193 shops with, on average, 7 employees per shop. Employees are centrally hired and assigned to shops. They do not move between shops; hence, they cannot sort into treatment, in contrast to Lazear (2000) in which half of the productivity effects of an individual bonus was owing to selection. Our randomization also accommodates the second identification issue, as we compare units of the same firm that all use the same technology and operate under similar conditions.

Regarding the realism of the experiment, it is undertaken in a firm that has been established for many decades and which still exists. The employees do not know that they are part of an experiment,[2] and carry out their normal day-to-day job. The only intervention is a team bonus of up to €300 per month, conditioned on preexisting

[1] To avoid confusion, our research question is not whether the joint introduction of team organization *and* team incentives increases performance, but whether a team bonus, given an existing team technology, leads to economically significant efficiency gains. Also, our paper is about monetary incentives offered to teams depending on their absolute performance. It is different from other field experimental studies that focus on the salience of existing incentive schemes (Englmaier, Roider, and Sunde forthcoming), on relative performance evaluation between individuals (Barankay 2012) or teams (Lavy 2002; Delfgaauw et al. 2013; Delfgaauw et al. 2014), and lab experiments on incentives (Nalbantian and Schotter 1997; Kocher, Strauß, and Sutter 2006).

[2] Except for the project team in management and the workers' council, no one was aware of our involvement in the implementation of the team bonus, and management took care of all communications. The firm used the term "pilot," a term it often employs when introducing new practices for a limited period of time.

sales targets.[3] In particular, we do not change the organization of the workplace: the shops continue to operate under a system of teamwork in which workers carry out a variety of tasks, such as handling the goods delivered, operating the oven, or serving customers.

We find that the team bonus increases sales in the treated shops by around 3 percent, which is equivalent to one-third of the sales standard deviation. Wages increase by 2.2 percent, on average, and up to 12 percent for some employees. 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 treatment effect is stable over the entire treatment period (April to June 2014). Contamination and gaming of the incentive scheme appear to play no role. Many of the shops in the treatment group increase their sales beyond the level at which the bonus was capped, which indicates potential efficiency gains of simple team bonus schemes beyond the ones we observe. Because the bonus was profitable for the firm, the management decided to roll out the scheme to all of their shops. Over the course of six additional months, we observe that the treatment and control shops' performance converged to each other, suggesting long-term stability of the treatment effect. The profit margin is estimated to increase by more than 60 percent after the rollout.

An average effect of 3 percent seems rather small in absolute terms, but a number of things should be noted. First, the elasticity of our bonus scheme is 3.8 (i.e., an increase of the costs by 10 percent results in an increase of sales by 38 percent) which is orders of magnitude higher than the elasticity of many marketing practices: for example, in their meta-analysis, Albers, Mantrala, and Sridhar (2010) find that the elasticity of resources invested in "personal selling" (sales persons selling products to customers in personal meetings) is 0.34; the average sales-to-advertising elasticity is estimated as 0.1 (Tellis 2004).[4] Second, we changed only one HR management practice rather than the entire HR system (as in Ichniowski, Shaw, and Prennushi 1997). Third, Germany is a country with high levels of managerial efficiency and product market competition, leading to high productivity, especially in retail,[5] suggesting less scope for improvement through an experiment like ours compared to countries with lower efficiency levels, such as India (Bloom et al. 2013). Finally, in contrast to Lazear (2000), the output effect is entirely driven by incentives (the experiment shut down the selection channel).

[3] Individual bonuses are infeasible for technological reasons. Objective measures of performance are available on the level of shop teams only, and subjective evaluations are hard to implement. The firm had actually experimented with shop supervisor bonuses conditioned on the subjective evaluations of mystery shoppers. The bonus was discontinued because evaluations had low interpersonal reliability, the subject of a companion research project of ours.

[4] 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 percent per month (probably an overestimate because of nonrandom selection). With the costs of redesign of at least €150,000 per shop, the historical share of value added in output of 0.56, the German corporate tax rate of 30 percent (needed to calculate tax rebate), and a liberal lending interest rate of 3 percent per year, the average return on investment over a ten-year horizon would be less than 0.6 percent per year.

[5] According to Bloom and Van Reenen (2007) and Bloom et al. (2012), German manufacturing firms, hospitals, and schools have managerial efficiency levels that are among the highest in the world, and Baily and Solow (2001) show that German retail firms have one of the highest levels of productivity worldwide. The German retail market is 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.

What is the mechanism behind the treatment effect? Notice first that there was no increase in employment in the treatment shops. Hence, sales teams must have exerted more effort. This additional effort did not, however, result in higher average sales per customer; rather, the increase in sales is commensurate with the increase in customer visits. It is likely that the extra effort was allocated into dealing more efficiently with the flow of incoming customers. In line with this is the fact that in towns with more than 100,000 inhabitants, the sales increase is 7.7 percent, but in the country-side, the effect is zero. In high population-density areas, demand is more likely to be concentrated around certain hours, say lunch, providing the employees with an opportunity to increase effort in productive ways, namely to deal with queues efficiently in order not to lose customers. In rural areas, this opportunity is not available. We discuss this explanation in Section V, by using additional data and what we learned from interviews with the employees.

The above mechanism is in line with a simple agency model (see Section IV) predicting that the effect of the bonus crucially depends on the productivity of agents' efforts. The model also generates other predictions under what conditions the team incentive is more likely to work, and does a good job in explaining heterogeneous treatment effects.[6]

One prediction, and the associated empirical result, are particularly noteworthy. For legal reasons, not all workers could be incentivized: the so-called "mini-jobbers"[7] who represent around 28 percent of the headcount had to be excluded from the bonus. This institutional specificity provides a source of exogenous variation in the share of non-incentivized workers in a team, at a given bonus size. The model predicts that the treatment effect decreases with the proportion of work hours provided by non-incentivized workers. This prediction is confirmed by the data, suggesting important complementarities between the members of the teams. Furthermore, peer pressure (Kandel and Lazear 1992; Mas and Moretti 2009) appears to have its limits: the incentivized team members did not succeed in putting pressure on their (non-incentivized) teammates to work harder.

The other predictions of the model and the associated empirical results provide some guidance for the application of team incentives. We find that a team bonus works better for employees with lower costs of effort (empirically, younger workers), and in teams that historically have been underperforming.

Finally, while the welfare-enhancing effects of the bonus on profits and wages must in principle be set against possible welfare costs caused by higher stress levels that accompany monetary incentives (Cadsby, Song, and Tapon 2007), we find no adverse effects of our bonus scheme on job or life satisfaction, or organizational commitment, all of which we monitored in our own firm-wide employee survey. Neither, do the quit rates seem to be affected by the treatment. Thus, we are confident that our bonus scheme is a "win-win" for the firm and for the workers, in line

[6] We do not want to deny that behavioral forces (as analyzed most prominently by Kandel and Lazear 1992, but also Mohnen, Pokorny, and Sliwka 2008; Burks, Carpenter, and Goette 2009; Friebel and Schnedler 2011; von Siemens and Kosfeld 2014) may strengthen or weaken the effect of the team incentive on effort choice, but do believe that our rather parsimonious agency model generates a number of interesting, testable predictions in line with the data.

[7] Mini-jobbers are allowed to earn up to €450 per month, often in addition to receiving unemployment benefits. Beyond that threshold they are fully taxable. Tazhitdinova (2015) investigates mini-jobbers' labor supply in Germany, and provides interesting background on this labor market institution.

with the decision to roll out the bonus to all workers, taken by management and supported by the workers' council.

Concluding that our bonus scheme is a viable "investment in people" project, we also believe our results to be widely applicable: retail is one of the largest sectors in the world in terms of employment.[8] Many firms in the global economy employ similar types of teamwork (for instance, in catering, airlines, or hotels), and our bonus is simple and relatively easy to implement. It is important to stress that our teams are relatively small, so that free-riding is less of an issue than in large units such as divisions of corporations, but so are many of the teams in service operations.

In what follows, we provide the information needed to understand the setting and goals of the experiment, and then in Section II discuss the details of our intervention. Section III discusses the research design, in particular, the choice of an appropriate estimator. Section IV introduces our agency model of teamwork and, in Section V, we provide predictions and empirical results in line with them. Section VI summarizes the effect of the bonus on the firm and what can be learned from the rollout. Section VII looks at the effect on the workers. In Section VIII, we argue that the results are robust against a number of concerns, before highlighting some implications in the concluding remarks.

I. Background

A. *The Study Firm and the Challenges Faced*

The firm entails 193 bakery shops with a total of 1,300 employees. Like many of its competitors, the chain had developed its business model since 1980, by exploiting the benefits of attractive locations and economies of scale. In 2011, however, discount retailers Aldi and Lidl began to sell freshly-baked bread and related products in their dense network of existing shops, with significant success. Their bread is considered to be of similar quality to that of the chains, but is sold at much lower prices, which forces the incumbent chains to rethink their strategy. Our study firm moved into the market for snacks, cakes, sandwiches, and beverages traditionally covered by cafés and fast-food chains. Substantial investments in shop design were carried out and additional marketing instruments introduced, some of them in pilot studies. As a complement to the strategic shift into more service-oriented product lines, HR practices were reconsidered with the goal of motivating employees to engage more actively with their customers. After intensifying training and experimenting unsuccessfully with hiring more qualified employees to try to improve customer service, the firm approached us for help.

B. *HR Management Practices*

The firm has a well-defined hierarchical management structure, at the top of which are the general and district managers who oversee the shops. Shops are managed by supervisors who usually work full-time and ensure the efficient deployment

[8] In Germany, more than 3 million people (7 percent of the labor force) work in retail, and in the United States the figure is 14.9 million (10.2 percent of the labor force).

of workers in the shops, as well as compliance with technological and accounting procedures. Shop supervisors do not have a say in strategic matters (e.g., product mix, shop concept, prices, advertising campaigns, etc.) or personnel policies such as hiring, workforce composition, allocation, and incentives.

Prior to the experiment, the firm paid incentives to its managers and shop supervisors but not to regular sales agents. For middle managers and shop supervisors there is a detailed, centrally-managed system of key performance indicators (KPIs), against which they are evaluated and paid. For district managers (each overseeing 10 to 15 shops), the KPIs consist of sales, personnel costs, and customer service evaluations obtained from monthly "mystery shopper" visits. Shop supervisors have similar KPIs, except that these are based on the performance of their shops alone. Sales are by far the most important KPI for managers and supervisors. There is a step-wise bonus that depends on exceeding a predetermined sales target that cannot be renegotiated during the course of the respective year. Sales targets are determined at the end of the preceding year, based on past sales and a correction for the general trend in sales (minus 2 percent in 2014). In the data there is a correlation of 97 percent between the actual target and the target as predicted by this rule.

C. *Teamwork*

An average shop employs a team of 7 employees (4 full-time equivalents), average monthly sales are around €28,000, and there are around 10,000 customer visits per month (see column 1 in Table 1). In a typical shop, the workers carry out a variety of interconnected and often simultaneous tasks, such as handling goods, operating the oven, serving customers, etc. The volatility of demand makes task specialization expensive (Friebel and Yilmaz 2017), as workers would be idle much of their time. Instead, the workers are expected to help each other. For instance, consider a worker in a not-so-busy shift who can prepare sandwiches for the colleague(s) in a busier shift or do nothing. Or, consider a worker who could clean the oven or help a colleague deal with a queue of customers. Workers helping each other, 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 rise to team organization and, if anything, team-based compensation. Indeed, detailed interviews carried out in several randomly picked shops in December 2014 revealed that employees perceived their work as teamwork. Furthermore, there is only one cash register per shop, creating substantial congestion in peak times and thus requiring seamless coordination to serve customers quickly.

D. *Forms of Employment and the Mini-Jobbers*

Regular sales agents make up about 55 percent of the sales force and they are predominantly unskilled. Prior to the experiment, they received fixed wages, determined by collective agreements on the industry level (€9–€11 per hour, depending on tenure). Roughly 85 percent of them are part-time employees, who, on average, work 26.5 hours per week. All of them pay income tax, and around two-thirds of them have permanent contracts. Twenty-eight percent of the sales force are mini-jobbers who work on average 10.7 hours per week and, often in addition to

Table 1—Pretreatment Shop Characteristics

	All shops ($N = 193$)	Control ($N = 96$)	Treatment ($N = 97$)	Diff: control versus treatment	t-test p-value
Panel A. Quantitative performance indicators					
Mean monthly sales	27,820	27,453	28,194[a]	741	0.695
	(13,094)	(11,481)	(14,542)	(1,890)	
Mean monthly sales (in logs)	10.15	10.14	10.15	0.01	0.846
	(0.40)	(0.39)	(0.41)	(0.06)	
Sales trends (year-on-years sales growth)	−0.04	−0.04	−0.03	0.01	0.322
	(0.13)	(0.13)	(0.12)	(0.01)	
Unsold goods as percent of sales	16.04	16.16	15.92	−0.24	0.694
	(6.80)	(7.01)	(6.90)	(0.60)	
Mean number of customer visits	10,079	10,028	10,131	103	0.856
	(3,969)	(3,921)	(4,018)	(566)	
Frequency of achieving the sales target	35.5%	35.8%	35.2%	−0.6%	0.860
Panel B. Qualitative performance indicators					
Mean mystery shopping score	97.6%	97.6%	97.6%	0.0%	0.826
Panel C. Locations of shops					
Big town	30.0%	33.3%	26.8%	−6.5%	0.194
Mean monthly property rent per m²	8.79	8.94	8.72	−0.22	0.432
(zip code)	(1.84)	(1.85)	(1.87)	(0.27)	
Total number of other bakeries	2.67	2.81	2.52	−0.30	0.462
(1 km radius)	(2.75)	(2.88)	(2.60)	(0.40)	
Total number of Aldi/Lidl shops	0.89	0.89	0.90	0.01	0.932
(1 km radius)	(0.90)	(0.92)	(0.87)	(0.13)	

Notes: Standard deviations are in parentheses. Column 5 reports the p-values of the two-sided t-test of equality of the means. Panels A and B: the data are from January 2012 to March 2014. We drop a few shop-month observations as some shops were closed for several weeks because of refurbishments. Mystery shopping score: the scale is 0 percent to 100 percent. Panel C: the data are from March 2014. Big town refers to municipalities with more than 100,000 inhabitants. Property rent: the data are from *ImmobilienScout*24. The table shows the weighted average commercial and residential rent; we drop seven shops as we do not have the commercial property rents for one municipality.

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

receiving welfare benefits, earn up to €450 per month tax-free. (For additional information, see column 1 in Table 2). It is an interesting specificity of our setting that these mini-jobbers, because of their specific tax status, cannot earn additional money and hence were excluded from the benefits of the bonus scheme. We will use this institutional fact in our analysis to learn more about the anatomy of teamwork.

E. *Proposed Changes and Initial Concerns*

The company operates a well-functioning system of performance measurement. The lack of incentives for sales assistants, however, was striking. We (the researchers) converged quickly on the idea of implementing a team bonus and, in late February 2014, suggested that the management of the firm should implement a bonus for shop sales teams, including the shop supervisors, conditional on reaching or exceeding the sales targets.

One member of the management team remarked that "bonuses to sales staff were never on our agenda." Other members of the management team had considered a team bonus previously, but thought it would be ineffective because of the problem

Table 2—Pretreatment HR Characteristics

	All shops ($N = 193$)	Control ($N = 96$)	Treatment ($N = 97$)	Diff: control versus treatment	*t*-test *p*-value
Panel A. Characteristics of shop supervisors					
Mean weekly working hours	34.3	34.5	34.1	−0.4	0.532
	(4.3)	(4.0)	(4.5)	(0.6)	
Mean age, years	41.4	40.9	41.9	0.9	0.534
Share of females	94.7%	98.6%	90.7%	−7.9%	0.012
Mean monthly quit rate	0.7%	0.6%	0.7%	0.1%	0.691
Panel B. Characteristics of regular sales agents (excluding mini-jobbers)					
Mean number per shop	3.7	3.7	3.7	−0.1	0.770
	(2.0)	(2.0)	(1.9)	(0.3)	
Mean weekly working hours	26.5	26.7	26.2	−0.5	0.217
	(5.1)	(5.1)	(5.1)	(0.4)	
Mean age, years	40.5	40.3	40.6	0.3	0.768
Share of females	93.1%	92.5%	93.7%	1.3%	0.410
Mean monthly quit rate	2.5%	2.6%	2.4%	0.2%	0.458
Share of full-time agents (>35 hrs/week)	14.8%	14.7%	14.8%	0.1%	0.965
Share of agents with permanent contract	65.2%	64.0%	66.3%	2.3%	0.510
Share of agents without professional training in retail	63.5%	66.3%	60.8%	−5.5%	0.131
Panel C. Characteristics of mini-jobbers					
Mean number per shop	1.9	1.8	2.0	−0.2	0.389
	(1.3)	(1.1)	(1.4)	(0.2)	
Share in total, FTE-adjusted	12.0%	12.1%	12.0%	−0.1%	0.539
	(9%)	(9%)	(10%)	(1%)	
Mean weekly working hours	10.7	10.4	10.8	0.4	0.238
	(3.1)	(3.2)	(3.0)	(0.3)	
Mean age, years	31.3	31.1	31.6	0.5	0.695
Share of females	89.5%	90.0%	89.0%	−1.0%	0.580
Mean monthly quit rate	8.1%	8.2%	8.0%	−0.2%	0.732
Share of mini-jobber with permanent contract	28.5%	29.0%	28.0%	1.0%	0.848
Share of mini-jobber without professional training in retail	99.5%	100.0%	99.0%	0.8%	0.599
Panel D. Employee attitudes					
Mean commitment score	4.46	4.50	4.42	−0.08	0.523
	(1.62)	(1.55)	(1.69)	(0.12)	
Mean job satisfaction score	4.39	4.45	4.33	−0.11	0.422
	(1.54)	(1.51)	(1.57)	(0.14)	
Mean overall satisfaction score	4.94	4.98	4.90	−0.08	0.548
	(1.66)	(1.63)	(1.70)	(0.14)	

Notes: Standard deviations are in parentheses. Column 5 reports the *p*-values of the two-sided *t*-test of equality of the means. Panels A–C are based on the personnel records from the firm, excluding apprentices and interns (18 in the control and 11 in the treatment group). All data are from March 2014, except for the mean monthly quit rate (data are from January 2012 to March 2014). Panel D reports the means of the job 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. Response rate in the survey: 80 percent.

of individual free-riding. Some senior managers were afraid that bonus payments could prove to be a financial burden on the firm. In particular, the bonus would need to be paid even to those shop teams that would have reached their sales targets in any case.

These concerns relate to one of the most important questions in the literature on management practices: Why do some firms adopt productivity-enhancing management practices while others, even though in the same industry, do not? The literature comes up with several reasons. In particular: (i) differences in product-market competition (Bloom and Van Reenen 2010; Syverson 2011; Bloom et al. 2014); (ii) lack of knowledge (Bloom at al. 2013); (iii) organizational capabilities (Bandiera, Barankay, and Rasul 2011; Ichniowski and Shaw 2012).

Our experiment addresses all these points. Because of intensified product-market competition, the firm decided to fundamentally rethink its HR management practices. A lack of knowledge and awareness had prevented the firm from adopting sales-staff incentives sooner. There were several limitations on the resources the firm's managers could commit to new projects, given their existing responsibilities. The HR personnel, for instance, would need to spend time on administering the bonus without directly benefiting from it. Tensions between new and existing management practices caused additional resistance: the team bonus would imply higher personnel costs, whereas the sales benefits were not clear from the outset. As district managers' bonuses depend on both sales and personnel costs, they were skeptical.

F. *Getting to an Agreement*

It is interesting to note that it was relatively straightforward to deal with the mild forms of organizational resistance described above. In particular, we took care of a substantial part of the administrative work related to the implementation of the bonus. We ran simulations of the bonus effects on sales and personnel costs, showing that the team bonus payments were likely to be lower than €20,000 per month when half of the shops were treated and the monthly shop bonus was capped at €300. To deal with district managers' concerns about their bonuses, senior management decided that bonus payments to sales staff would be paid from a separate budget in order not to affect the district managers' cost KPI. District managers were quick to realize that they were likely to benefit from increased sales in their shops. The workers' council was also in favor of the bonus, because it was designed as a pure add-on payment and was a result of the high level of trust between the council and management.

To further gain trust we built up a record of research engagement with the firm prior to the experiment (see also List 2011). We achieved an "early success" by showing the company the limitations of their existing subjective performance evaluation system. We also reinforced trust through constant communication with managers at all levels of hierarchy. We received the shops' sales, financial and accounting, geographical, compensation, and personnel data since January 2012, which allowed us to conduct a very precise randomization procedure (explained in detail in Section III). We offered our advice free of charge and covered most of the research costs. The company provided the data and administrative support needed. Our main interfaces were the CEO, HR, the sales director, and a small group of district managers. The support of the workers' council turned out to be crucial. It assured legitimacy and commitment for the bonus, suggesting that institutions that one might have expected to be obstructive to change and experimentation, when convinced, will actually assist the experimenter by boosting trust and legitimacy within the firm.

II. Experimental Procedures

A. *Employee Survey*

Preparations for the experiment began by planning two waves of an employee survey. The first was in March 2014 (a month prior to the introduction of the team bonus), and the second was at the end of May 2014, in the middle of the treatment period.

The main variables measured in both waves of the survey were "satisfaction with the job context" and "overall satisfaction" (constructed by Hackman and Oldham 1980, and translated into German by van Dick et al. 2001), and "organizational commitment" (following Allen and Meyer 1990). The second survey also collected some additional data that were 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 resulted in response rates of 80 percent for the first and 60 percent for the second wave of the survey.

We conducted the survey for three reasons: (i) to check whether treatment and control samples are balanced with respect to employee attitudes; (ii) to see whether there is a treatment effect on employee attitudes; (iii) to test whether baseline attitudes affect the response to our treatment. The answer to the first point can be found in Section III; for the latter two points, we refer to Section VII.

B. *Information and Training about the Bonus Scheme*

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. We ensured that employees would not perceive themselves as being part of an experiment. Management handled all communications. Logos of our universities did not appear on these materials, and there was no mention of our research team in any communication about the bonus. Apart from senior 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 25, 2014, we instructed all of the district managers 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 in turn inform their shop's employees. We also instructed the managers on how to respond to questions about the bonus from the employees in the control-group shops, as follows: "This is a pilot. Every shop had the same chance to be drawn into the bonus scheme. The workers' council agreed to this procedure." The workers' council suggested that this response would be acceptable for the employees in control shops. 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. In general, we find no evidence for contamination (a point further discussed in Section VIII).

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

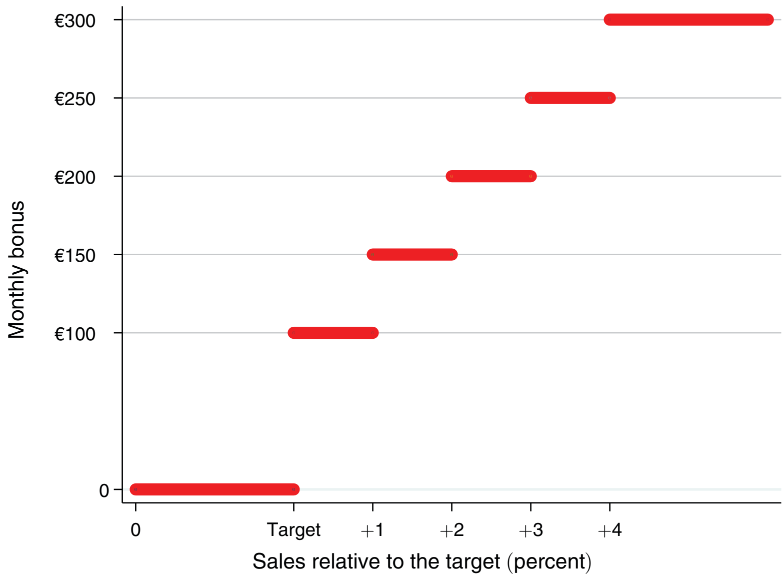

Figure 1. The Team Bonus

Notes: This figure illustrates the amount of bonus a shop sales team would receive depending on reaching and exceeding its sales target in a given month. Not reaching the target brings no bonus. Reaching or exceeding the target by up to 1 percent awards a bonus of €100. Every percentage point on top of 1 percent above the target brings an additional €50 of bonus. The bonus is capped at €300 paid when the target is exceeded by 4 percent or more. The bonus is shared between the part-time and full-time employees in the shop (excluding mini-jobbers) in proportion to their working hours during that month.

scheme for 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 income below that level is tax-free. Therefore, providing 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.

C. *The Bonus Scheme*

Figure 1 illustrates the bonus scheme offered to the treatment shops. Shops that reach the sales target for the month received a bonus of €100, to be shared between the part-time and full-time employees (including the supervisors) in the shop in proportion to their working hours during that month. The bonus increased by €50 for each percentage point above the target and was capped at €300 per month for exceeding the target by 4 percent or more. Hence, the team in a shop could make additional 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 example, a 1 percent increase above the sales target for a medium-sized shop would be equivalent to selling ten additional rolls, two loaves of bread, two sandwiches, and four cups of coffee per day.

This bonus scheme may be criticized on theoretical grounds for being susceptible to the strategic behavior of employees around the bonus cutoffs. However, in designing an incentive scheme one faces the trade-off between optimality versus clarity, and verifiability and approval of the scheme by its stakeholders. Our bonus scheme reflects this trade-off. In fact, it is not specific to our study environment since "threshold" bonuses are widespread (see Section IX). We do nevertheless address the possibility of "gaming" in Section VIII.

III. Research Design

In Section IV, we present a simple agency model to organize our thoughts both about the expected treatment effect of the bonus on shop-level sales, and treatment heterogeneities. We here would like to stress that our study firm has many advantages for this type of research. Management gave us access to detailed data about a large number of shops that use the same technology, and the data span a long period of observation. The experiment is likely to generate evidence that is broadly applicable because its competitive and work environment is comparable to many other retail and service firms in the global economy.

An important element of our research design is the choice of the appropriate estimator for our experiment. Frison and Pocock (1992), in medical research, and McKenzie (2012), in development economics, discuss three estimators that could be applicable in our setting: POST (single-difference estimator), CHANGE (difference-in-differences estimator), and ANCOVA (POST controlling for pretreatment average outcome).

The POST estimator is the following single-difference estimator:

$$\ln\left(sales_{it}\right) = \beta \cdot treatment_i + month\ fixed\ effect_t + controls_{it} + error_{it}, \tag{1}$$

where $\ln\left(sales_{it}\right)$ is the log sales in shop i and month t, $controls_{it}$ are time-variant variables on the shop-level (log total hours worked and dummies for shop refurbishment); $error_{it}$ is the idiosyncratic error term which is clustered at the shop level.

The CHANGE estimator is the following difference-in-differences estimator:

$$\begin{aligned} \ln\left(sales_{it}\right) = {} & \beta \cdot treatment_i \cdot after_t + month\ fixed\ effect_t \\ & + shop\ fixed\ effect_i + controls_{it} + error_{it}. \end{aligned} \tag{2}$$

Here, $after_t$ is a dummy variable equal to one for all months from April to June 2014, and zero for all months from January 2012 to March 2014, i.e., in contrast to POST, CHANGE uses both pre- and post-treatment data.

Finally, ANCOVA estimates (for the observations from April to June 2014):

$$\begin{aligned} \ln\left(sales_{it}\right) = {} & \beta \cdot treatment_i + month\ fixed\ effect_t + \delta \cdot \overline{\ln\left(sales_{i,PRE}\right)} \\ & + controls_{it} + error_{it}, \end{aligned} \tag{3}$$

where $\overline{\ln\left(sales_{i,PRE}\right)}$ is the average of the log sales in the pretreatment period (January 2012 to March 2014).

According to McKenzie (2012) and provided randomization is successful, all three estimators will give an unbiased estimate of the average treatment effect (β).

Whether CHANGE or ANCOVA, versus POST should be employed for estimating the treatment effect depends on the estimators' efficiency benefits and costs. The benefits stem from using pretreatment observations to separate treatment effects from noise (intuitively, the more rounds of observation, the more precisely the group means are estimated), while the costs are owing to extra parameters that need to be estimated (fixed effects in CHANGE and δ in ANCOVA). McKenzie (2012) shows that CHANGE is more efficient than POST when the autocorrelation in the outcome variable is greater than $1/(1 +$ the number of pretreatment observation periods), and that ANCOVA is the most efficient estimator of the three (unless the autocorrelation is zero, in which case ANCOVA and POST are equally efficient). In our data, autocorrelation in sales is in excess of 0.95 and there are 27 pretreatment months. Consequently, we use ANCOVA in all regressions (but will also report CHANGE in our main regression table).

To minimize the standard errors in the estimations, we follow Barrios' (2014) randomization strategy. We first run a regression of log sales on labor input with month and shop fixed effects for 2012 and 2013. Shops are then ranked according to the predicted sales and randomized within the pairs of shops with adjacent ranks, except for the median-ranked shop (#97) that was randomly assigned to the treatment group. The resulting treatment and control groups comprised 97 and 96 shops, respectively. Power calculations reveal that the sample size is more than sufficient: on the basis of 27 months of observations pretreatment (January 2012 to March 2014) and three months of observations post treatment (April to June 2014), we would need 70 shops in each group to detect a 3 percent treatment effect at a 5 percent significance level with the probability 0.9.

Randomization succeeds in generating treatment and control groups that are balanced in terms of pretreatment sales (our key outcome variable). Tables 1 and 2 (columns 2 to 5) show that treatment and control groups are also balanced in all other potentially relevant characteristics. Table 1 lists quantitative and qualitative shop performance indicators, and location characteristics, while Table 2 shows balance concerning the different subgroups of employees.

We would like to stress two particularly important facts. First, the mini-jobber characteristics are balanced (Table 2, panel C) making it possible to discuss heterogeneous treatment effects with respect to mini-job labor supply in the shops (see Section V). Second, the employee survey results are also balanced (Table 2, panel D), although the survey results were not yet available when we carried out the randomization.

IV. Agency Model

The simple agency model discussed below is adapted to the specific setting of our study firm. Members of a team decide individually what effort level to choose, given a threshold bonus like the one we used in our field experiment. In describing the model, we deliberately use references to the specific situation of our study firm in order to make the link between the model and the empirical analysis as clear as possible.

We consider a shop team of N workers; in our firm, these are on average, seven. The team produces sales y that depend on the team's total effort E, the productivity of team effort a, and additive noise v with a probability distribution function $\phi(v)$ symmetric around zero:

$$y = a \cdot E + v. \tag{4}$$

Here, total effort is a CES aggregate of individual efforts e_i, $i = 1, \ldots, N$:

$$E(e_1, \ldots, e_N) = \left(\sum_{i=1}^{N} e_i^{\rho}\right)^{\frac{1}{\rho}}, \tag{5}$$

where, following Alchian and Demsetz's (1972) definition of team production, we assume that individual efforts are complementary, that is, $\rho < 1$.[9] We will later also discuss the empirical and anecdotal evidence in favor of complementarities between team members.

The firm uses a team bonus $B > 0$, which is paid if sales exceed an exogenously given target y_0. The bonus is split evenly between the team members, excluding, for legal reasons, the mini-jobbers whose share we denote as θ.

To keep the complexity of the model to a minimum, we only consider one target rather than the multi-step bonus scheme implemented in our firm. The *expected* bonus is a function of the team effort E which increases the probability of producing sales above the predefined threshold:

$$g(E) = B \cdot \Pr(a \cdot E + v \geq y_0) = B\Phi(a \cdot E - y_0), \tag{6}$$

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

Given the above, incentivized team members choose, independently and simultaneously, effort levels e_i to maximize their expected individual payoffs,

$$\pi(e_i, e_{-i}) = w_0 + \frac{1}{N} B\Phi(a \cdot E - y_0) - b \cdot c(e_i), \tag{7}$$

where w_0 is a fixed wage, $c(e_i)$ is a monotonic, continuous, twice-differentiable, and convex cost of effort function, and b is a parameter measuring the difficulty of effort.

In line with the reality in our firm, the effort choice is constrained from below by a minimally acceptable level e_0, which stems either from some intrinsic motivation as in Holmström and Milgrom (1991), or from monitoring activity by supervisors and managers, as in Lazear (2000). We consider that e_0 is the same for all team members. Mini-jobbers who do not benefit from the bonus always carry out e_0.

[9] Similar predictions could be generated by imposing assumptions on the curvature of the costs of effort function, as in Itoh (1991), at the expense of more involved and less intuitive assumptions.

Solving the model, we find that the optimal individual effort level e^* must satisfy the following conditions:

$$(8) \qquad \left.\frac{d\pi}{de_i}\right|_{e_i=e^*} = aN^{\frac{1-2\rho}{\rho}} B\Phi'\left(aN^{\frac{1}{\rho}}e^* - y_0\right) - b \cdot c'\left(e^*\right) = 0,$$

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

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

implying $e^* > e_0$. Alternatively, when $\left.\frac{d\pi}{de_i}\right|_{e_i=e_0} \leq 0$, $e^* = e_0$.

V. Predictions of the Model and Empirical Results

The first prediction follows in a straightforward way. Leaving aside the issue of multiple equilibria,[10] the other predictions derive from comparative statics on the conditions in (8), under the assumption of a positive effort response to the given bonus B. While predictions VA, VD, and VE follow in a straightforward way from (8), the proofs to VB and VC are in Appendix A.

A. *The Effect of the Bonus on Sales*

The model tells us that a team bonus $B > 0$ will lead to increased effort and, hence, expected sales, provided the marginal benefit of effort given the bonus exceeds its marginal costs at the minimum acceptable level e_0. This implies that in the presence of managerial instruments other than team incentives, effort and sales in some teams may respond to the bonus more strongly than in others, and in some teams, sales may not respond at all.

Table 3, panel A, compares the quantitative performance indicators of treatment and control shops. Reflecting the secular downward trend in the bakery market, sales and the number of customer visits have gone down for both treatment and control shops, but significant differences are noticeable on both dimensions, suggesting a positive treatment effect.

Figure 2 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. Also instructive is to look at the treatment's effect on the sales rankings of shops within the treatment-control pairs that result from our assignment procedure:[11] Only 18 percent of the shops in

[10] Multiple symmetric equilibria are possible because the bonus rule may induce strategic complementarity (Cooper and John 1988).

[11] Recall that out of a pair of two shops with adjacent ranks in terms of predicted sales performance, one was randomly assigned to the treatment and the other to the control group.

TABLE 3—COMPARISON OF CONTROL AND TREATMENT GROUP IN THE TREATMENT PERIOD, APRIL–JUNE 2014

	Control ($N = 96$)	Treatment ($N = 97$)	Diff: control versus treatment	Diff-in-diff p-value
Panel A. Quantitative performance indicators				
Mean monthly sales	25,376	26,995	1,619	0.061
	(10,708)	(15,036)	(1,844)	
Mean monthly sales (in logs)	10.06	10.10	0.04	0.034
	(0.40)	(0.42)	(0.06)	
Unsold goods as percent of sales	22.88	22.35	−0.53	0.940
	(9.80)	(13.30)	(1.36)	
Mean number of customer visits	9,115	9,465	350	0.062
	(3,582)	(3,790)	(529)	
Frequency of achieving the sales target	44.8%	49.1%	4.3%	0.442
Panel B. Qualitative performance indicators				
Mean mystery shopping score	98.2%	97.6%	−0.6%	0.295
Panel C. Mean monthly quit rate				
Shop supervisors	1.5%	1.1%	−0.4%	0.493
Sales agents (excluding mini-jobbers)	1.7%	2.1%	0.4%	0.250
Mini-jobbers	5.1%	5.8%	0.7%	0.448
Panel D. Employee attitudes				
Mean commitment score	4.20	4.24	0.03	0.468
	(1.28)	(1.35)	(0.12)	
Mean job satisfaction score	4.39	4.48	0.08	0.245
	(1.34)	(1.20)	(0.14)	
Mean overall satisfaction score	3.59	3.72	0.13	0.162
	(1.12)	(1.02)	(0.09)	

Notes: Standard deviations are in parentheses. Column 5 reports the p-values of the two-sided significance test for the difference-in-differences estimate of the treatment effect. Panels A–C: Data are from April to June 2014. We drop three shop-month observations as shops were closed for several weeks because of refurbishments. Panel D: The second employee survey was administered at the end of May 2014 with a response rate of 60 percent.

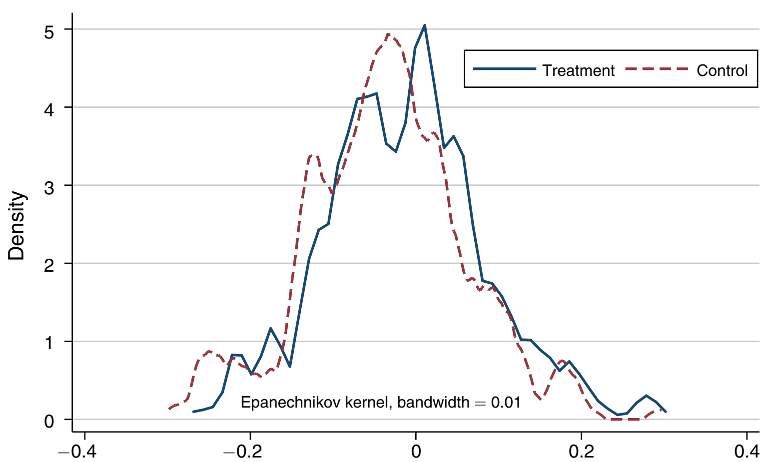

FIGURE 2. KERNEL DISTRIBUTION OF THE YEAR-ON-YEAR SALES GROWTH IN THE TREATMENT PERIOD, APRIL–JUNE 2014

Table 4—Average Treatment Effect, April–June 2014

	$\ln(sales_{it})$		
Estimators:	POST	CHANGE	ANCOVA
$Treatment_i$	0.029		0.032
	(0.034)		(0.014)
$Treatment_i \times after_t$		0.032	
		(0.013)	
Month fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Shop fixed effects	No	Yes	No
Average $\ln(sales_{it})$ pretreatment	No	No	Yes
Observations	576	4,916	576

Notes: The table shows the estimated treatment effects for our treatment period (April to June 2014). In all regressions, $\ln(sales_{it})$ in shop i and month t is the dependent variable. POST is a single-difference estimator, CHANGE is a difference-in-differences estimator, ANCOVA is a single-difference estimator controlling for pretreatment average $\ln(sales_{it})$. $Treatment_i$ is a dummy set to one for the treatment shops (zero otherwise); $after_i$ is a dummy set to one for the treatment period (April–June 2014). Controls are time-variant variables on the shop-level (log total hours worked and a dummy set to one if a shop was refurbished in the previous month (zero otherwise)). We drop three shop-month observations in our treatment period as shops were closed for several weeks because of refurbishments. All observations are from January 2012 to June 2014. Coefficient standard deviations clustered by shop are in parentheses.

the control group improved their "within-pair" ranking after the treatment, while in the treatment group it is 37 percent, a clear sign of a strong treatment effect.

Table 4 presents the results for POST, CHANGE, and ANCOVA. The average treatment effect is around 3 percent and is statistically significant for CHANGE and ANCOVA.[12] The point estimate of the treatment effect obtained with the POST estimator is similar as the ANOVA and CHANGE estimates, but it is statistically insignificant, reflecting the lower efficiency of the POST estimator (see our discussion in Section III). Calculating the treatment effect in each month, we find it to be 3.1 percent in April, 3.7 percent in May, and 3.3 percent in June 2014. This represents a steady effect without noticeable abatement.

All other predictions look into treatment heterogeneity, and use ANCOVA regressions with sample splits by quartiles. We summarize the estimations in Table 5, panels A–C. Notice that treatment and control groups are balanced in all the characteristics we analyze below.[13]

B. *The Effect of the Bonus Depending on the Share of Non-Incentivized Members in Shop Teams*

According to the model, the effect of the bonus on total effort decreases with θ, the share of non-incentivized team members. Empirically, we use the share of total

[12] Bootstrapping produces standard errors of similar magnitude.

[13] There are no unambiguous predictions on how team size N affects the impact of the bonus on team effort (see Appendix B). In regressions, we find some evidence that the effect of the bonus is stronger in larger teams.

Table 5—Treatment Effect Heterogeneity, April–June 2014

Panel A. Treatment effect by the shop-average share of work hours delivered by mini-jobbers

Quartile 1 (mean share: 2.4%)	Quartile 2 (mean share: 8.5%)	Quartile 3 (mean share: 13.4%)	Quartile 4 (mean share: 24.5%)
0.074	0.053	−0.001	−0.006
(0.035)	(0.028)	(0.023)	(0.025)

Panel B. Treatment effect by pretreatment deviation of sales targets

Panel B1. Distance measure: pretreatment average sales/target difference

Quartile 1 (mean deviation: −14.0%)	Quartile 2 (mean deviation: −6.5%)	Quartile 3 (mean deviation: −2.9%)	Quartile 4 (mean deviation: 2.5%)
0.055	0.041	0.047	0.001
(0.032)	(0.031)	(0.026)	(0.018)

Panel B2. Distance measure: pretreatment frequency of achieving the target

Quartile 1 (mean deviation: 9.8%)	Quartile 2 (mean deviation: 24.3%)	Quartile 3 (mean deviation: 41.7%)	Quartile 4 (mean deviation: 68.6%)
0.057	0.049	0.030	−0.012
(0.024)	(0.028)	(0.030)	(0.018)

Panel C. Treatment effect by shop-average employee age

Quartile 1 (mean age: 34 years)	Quartile 2 (mean age: 39 years)	Quartile 3 (mean age: 44 years)	Quartile 4 (mean age: 50 years)
0.059	0.076	−0.017	0.016
(0.029)	(0.029)	(0.028)	(0.020)

Notes: The table gives the estimated treatment effect in our treatment period (April to June 2014) for a given quartile of the variable in question. The regression specification is similar as in specification (2) in Table 4 (ANCOVA). We drop three shop-month observations as shops were closed for several weeks because of refurbishments. In each panel, we report in parentheses the mean of the stratifying variable within the respective quartile. Panels A, C, and D: split in quartiles is based on data from the treatment period (April to June 2014). Panel B: split in quartiles is based on data from the pretreatment period (January 2012 to March 2014). Coefficient standard deviations clustered by shop are in parentheses.

work hours carried out by mini-jobbers as measure for θ. This variable is orthogonal to our treatment (see Table 2, panel C), so we can directly test the prediction.

Panel A of Table 5 reports treatment effect estimates by quartiles of the shop-average share of work hours delivered by mini-job workers. We find that the treatment effect goes down with the share of work hours delivered by mini-jobbers, as predicted by the model. It is noteworthy that for the third and fourth quartile, the treatment effect drops to zero from 7.4 percent for the first, and 5.3 percent for the second quartile. This suggests that in teams in which a substantial part of the work is delivered by non-incentivized workers, the extra effort of the incentivized workers has no effect on team output. Put differently, in our shops, there seem to be complementarities between workers in general, and, more specifically, between mini-job and regular workers.[14]

We also interpret this finding as evidence for the limits of peer pressure. One may have thought that the incentivized members of the team would try to motivate or put pressure on the non-incentivized members; however, these measures (if actually present) are not enough to compensate for the lack of the mini-jobbers' effort.

[14] Iranzo, Schivardi, and Tosetti (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.

C. *The Effect of the Bonus Depending on the Distance between a Shop's Actual and Its Target Performance before the Treatment*

According to the model, effort under the bonus increases with the distance between a shop's actual performance and the target performance y_0 in the period before the bonus was introduced, and it drops to e_0 when the distance grows too large. This translates into the empirical prediction that in teams that have historically reached the sales targets less often, the bonus leads to larger effects on sales, 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 B of Table 5 reports treatment effect estimates by quartiles of historical distance to the sales target, measured as: (i) the difference between actual and target sales averaged for each shop over the pretreatment period (panel B1); and (ii) the frequency at which a shop achieves its target in the pretreatment period (panel B2). 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. Recall that sales targets for shops were adjusted from one year to another through a simple rule, a percentage increase or decrease depending on the year-on-year growth of the entire chain. While our results show that historically underperforming shop teams were successfully motivated to perform better, the mechanical target adjustment also means that the high-performing shops would not be motivated to perform even better. One could think of a more sophisticated way of determining sales targets in a history-dependent way for each shop. We are not in a position to judge whether the company decided not to do so because it might have feared ratchet effects (Laffont and Tirole 1988), or because of the limited organizational capacities it could rely on. Note finally that our regressions do not detect the existence of a threshold effect in contrast to what the model predicts.

D. *The Effect of the Bonus Depending on the Difficulty of Effort*

In the model, the effect of the bonus on total effort decreases in b. The proxy we use here is age, because memory, multitasking ability, and information-processing speed have been shown to decrease with age (de Zwart, Frings-Dresen, and van Dijk 1995; Salthouse 2010; Singh-Manoux et al. 2012). Hence, older workers may have larger marginal costs of effort, which translates into the prediction that in teams with older workers the bonus leads to smaller effects on sales.

Panel C of Table 5 reports treatment effects by quartiles of the shop-average workforce age. Shops with a younger workforce respond to treatment more strongly.[15] In our sample, younger workers are more likely to have low tenure (although the tenure/age correlation is low), to be mini-jobbers, to work part-time, and to have temporary (rather than permanent) job contracts. In regressions, none of these variables interact significantly with the treatment, but age does.[16]

[15] The results are also consistent with studies that found the mental processes supporting decision-making to change with age (Samanez-Larkin and Knutson 2015). In particular, younger people are more affected by monetary rewards (Rademacher et al. 2014).

[16] Average age may mask differences in a team's age distribution. We provide the frequency plots of the highest and the lowest quartiles in terms of average age in the shops in Appendix B.

E. *The Effect of the Bonus Depending on the Productivity of Effort*

In the model, total effort under the bonus treatment increases in a. We cannot test this prediction directly. Rather we would like to bring forward the argument that the productivity of effort and hence the scope of teams to increase sales is related to the potential prevalence of peak demand times and the possibility of reducing customers' waiting time.

In order to better explain what we have in mind, notice first, from the descriptive statistics, that the increase in sales is equivalent to the increase in the number of customers served, and that there is no significant increase in sales per customer. Hence, effort must have been productive when it was spent on dealing more effectively or increasing with customer flows (and not on increasing sales per customer).[17] But shop teams have little, if any, leeway to attract customers in the first place. However, they can work harder in order to reduce customers' waiting time by working faster and helping each other. This also increases the likelihood of customers returning. Interviews with sales agents and shop supervisors support this view.[18]

This mechanism relies on demand to have peak times that result in queues, and it will be more important when customers have higher opportunity costs of queuing, such as in populous, urban locations that have office workers who might come in for lunch, at the beginning, or the end of their work day. In contrast, in smaller towns and villages, peak times are less likely, and the opportunity cost of waiting is lower, leaving less scope to teams to increase sales by reducing waiting times. It should be noted that we do not have data that are grained finely enough to distinguish sales over a day, but that we only have monthly sales data.

Looking at the treatment effect by shop location, we indeed find that the treatment effect is largest, at 7.7 percent, in shops located in big towns ($>$ 100,000 inhabitants), going down to 2.6 percent in midsize towns, and zero in villages ($<$ 10,000 inhabitants). We believe this heterogeneity to be an important result that is in line with the model, which predicts that the bonus may have no effect at all, and that the magnitude of the effect depends on the productivity of effort, i.e., wiggle room of the team.[19]

To further examine the empirical support, we took the finer-grained data from *ImmobilienScout*24, the leading real estate platform in Germany. We use residential and commercial property prices per square meter in the 136 zip codes our shops are located in. Reflecting the idea that the opportunity costs of waiting in the queue are higher in more expensive areas, we interact our treatment effect with the

[17] We tested the possibility that increased friendliness was behind the treatment effect explicitly. Our research assistants went on a double-blind mystery shopping tour in May 2014 to evaluate shop staff friendliness on a Likert scale. We found no differences in friendliness between treatment and control group shops. We also found that the question "Would you like to have anything else?" was asked slightly more often in treatment shops in larger municipalities; whether the questions were asked or not played no role in explaining the increase in sales.

[18] We were told that within a shift, teammates' helping each other is crucial in peak times. Queues tend to be shorter when employees help each other. Helping the employee who is serving a customer, however, tends to be privately more costly than preparing bread for the oven or doing paperwork in the back office. Across shifts, employees can help the team to work faster by preparing the goods in quieter times to be sold at busier times in the next shift, so that more time could be spent on the customer then. Similarly, in off-peak times, employees can do maintenance rather than have a break, making more capacity available to their colleagues in peak times.

[19] The main qualitative results are the same if we use 50,000 inhabitants as the cutoff between big and midsize towns (treatment effects: 5.3 percent in big towns, 4.0 percent in midsize towns).

Table 6—Treatment Effect by Average Zip Code Property Price, April–June 2014

	(1)	(2)	(3)
Treatment effect (TE) at mean property price	0.036 (0.014)	0.031 (0.014)	0.035 (0.014)
TE × average per square meter price: commercial	0.037 (0.020)		
TE × average per square meter price: residential		0.036 (0.014)	
TE × weighted average by square meters			0.035 (0.015)

Notes: The table shows the estimated treatment effects for our treatment period (April to June 2014), interacted with the property rent (in March 2014, data are from *ImmobilienScout*24). The regression specification is similar as in specification (2) in Table 4 (ANCOVA). All the variables interacting with the treatment effect are standardized to have mean zero and standard deviation one. Thus, for example, the coefficient 0.037 on the interaction between the treatment effect and the average rental price of commercial property in specification (1) implies that a one-standard-deviation increase in commercial rental price is associated with a 0.037 higher treatment effect. We drop three shop-month observations as shops were closed for several weeks because of refurbishments. Column 1 and 3: we drop seven shops, as we do not have the commercial property rents for one municipality. Coefficient standard deviations clustered by shop are in parentheses.

average rental price for commercial and private real estate. The ANCOVA estimation results, reported in Table 6, are in line with the expectations. A one-standard-deviation increase in rental prices drives the treatment effect up by about 3 percent from the average of 3 percent.

VI. The Effect of the Bonus on the Firm and the Rollout

The average treatment effect on sales of 3 percent implies an extra €820 ($= [\exp(0.03) - 1] \times$ €27,000) of sales per month on average, or €238,620 (= €820 × 3 months × 97 shops) in all treatment shops over the treatment period. Given the historical share of value added in sales of 0.56, the implied value added gain is €460 and a total of €133,627.

The total team bonus payments in April to June 2014 amounted to €35,150, or 2.2 percent of the total labor costs in the treatment shops. There was a knock-on effect on shop supervisor mean bonuses of €240 per treatment shop per quarter, which is equivalent to an amount of €23,280 for all 97 treatment shops. District and senior manager bonuses increased by an estimated €4,500. We estimate the one-off costs associated with the implementation of the bonus scheme (i.e., printing and delivering materials, administrative support such as bonus calculations, etc.) and the costs of managers' time[20] at €25,000.[21] The total costs add up to €87,930. The net benefit from the scheme is hence €45,700 for the treatment shops in the treatment period.

We show below that workers also benefited substantially. Given this win-win situation, it is not surprising that the management decided to roll out the bonus scheme to all shops. The decision to do so was taken in the second half of June

[20] While we were not paid for our advice, one could consider the costs of our time up to implementation. Evaluated at the daily rate of consulting firms, the project's break-even would be reached in less than a quarter.

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

TABLE 7—TREATMENT EFFECT BY MONTH

Treatment period			Post-treatment period					
April	May	June	July	August	September	October	November	December
0.031	0.037	0.033	0.036	0.014	0.003	0.010	0.005	−0.002
(0.012)	(0.012)	(0.014)	(0.013)	(0.016)	(0.016)	(0.017)	(0.021)	(0.025)

Notes: The table shows the estimated treatment effects in each month (April–December 2014). The regression specification is similar as in specification (2) in Table 4 (ANCOVA). We drop several shop-month observations as shops were closed (either completely or for several weeks because of refurbishments). Coefficient standard deviations clustered by shop are in parentheses.

2014, communicated swiftly, and shops were informed in the course of the last days of June and first weeks of July 2014. We have data until the end of December 2014, i.e., for six additional months. We estimate an augmented regression equation that allows for treatment effects during the period of treatment and the rollout period. The results are shown in Table 7. We find that in the post-treatment period (July to December 2014), there are no statistically significant treatment versus control differences, except for the month of July, during which the rollout was communicated to all shops (and hence there still persisted differences for a subset of the shops).

Analogously to the analysis in Section V, we also investigated the within-pair ranking of treatment and control group shops in the rollout period, and find no effect at all (33 percent of the control, and 37 percent of the treatment group shops improved their "within-pair" ranking), a stark contrast to the significant treatment/control differences during the treatment period. The fact that we find no detectable "treatment effect" once the team bonus was rolled out supports our view that the introduction of the team bonus is causal for the increase of sales in the treated shop.

Projecting the cost and benefit calculations to the rollout period, the net gain becomes €140,000 per quarter for the entire chain. According to these calculations, each dollar spent on the bonus brings $3.8 of extra sales, or $2.1 of extra operational profit. Profit margins after the rollout cannot be computed precisely because we are lacking information about rental costs and depreciations. However, we can do a back-of-the-envelope calculation. According to top management, the firm's profit margin after taxes is usually below 1 percent (= €270 per shop per month). Our bonus scheme increased the net profits after the rollout in each store by around €240, leading to an increase of the after tax profit margin by more than 60 percent.

VII. The Effect of the Bonus on Workers: Wages, Inequality, Attitudes, Personnel Turnover

As depicted in Figure 3, close to 40 percent of the workers in the treatment group received a bonus at least once in the treatment period. For those who did receive a bonus, it averaged, for the treatment quarter, at €114.7 or 4.1 percent (SD: 2.3 percent) of the average recipient's quarterly earnings, with a median of 3.9 percent. We believe that this is a significant increase, not only because the employees are at the lower end of the wage income distribution. Rather, it should be noticed that this exceeds the pay workers would receive for an extra half workday, and that the other main source of wage increases because of the collective payment agreements are usually below 2 percent.

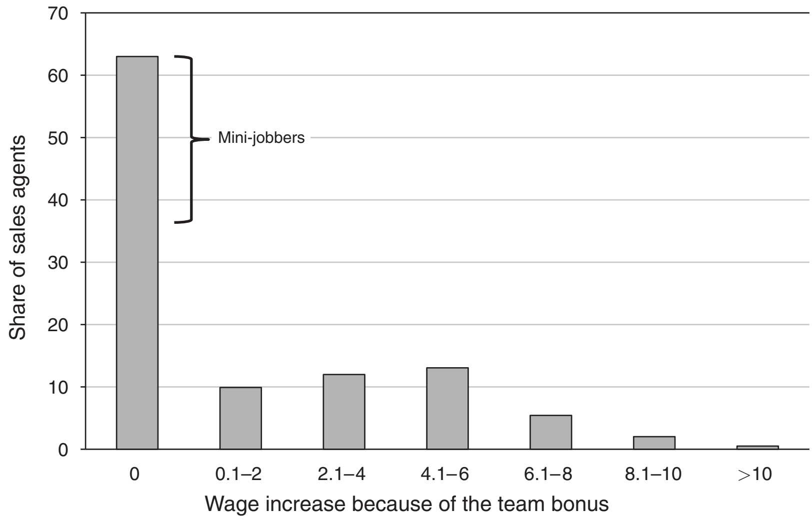

Figure 3. Distribution of Wages Increases Because of the Team Bonus in the Treatment Group, April–June 2014 (*Percent*)

Notes: "Sales agents" are defined as shop supervisors, regular sales agents, and mini-jobbers. The numbers do not include the knock-on effect that the treatment effect had on the shop supervisor bonus.

Lemieux, MacLeod, and Parent (2009) show that performance pay has affected wage inequality in the United States (the incidence of performance pay can explain one-fifth of the growth in the variance of (male) pay between the late 1970s and the early 1990s). Figure 3 shows that wage inequality in our firm has increased because of the bonus. Mini-jobbers (more about them below) constitute a substantial part of those employees who did not receive a bonus, but around 50 percent of the regular employees (including shop supervisors) also received no team bonus at all. There is also some inequality at the intensive margin. The wages of the top 25 percent of the bonus recipients increased by at least 5.5 percent, and one sales agent increased her wage by 12 percent.

We would argue that these are moderate levels of inequality, and that, because most of this inequality is across teams and not within, it is unlikely to cause problems in the work morale of employees. However, a potential problem is the fact that the mini-jobbers are excluded for regulatory and tax reasons. There could be multiple ways mini-jobbers could react to what may be perceived as procedural and distributional injustice. Arguably, one should be able to pick this up in the two waves of an employee survey we carried out before and during the treatment period. We measure whether average attitudes such as commitment, overall satisfaction, and job satisfaction were affected by the treatment. We find no treatment effect (see Table 3, (panel D) for the descriptive statistics, and Figures A2–A4 in Appendix B for the distributions).[22] Because of anonymity, we cannot break down the average effects by mini-jobber versus regular employee. It should also be noted that the second

[22] We also find no significant interaction effect between baseline attitudes and treatment on sales.

wave of the survey had a considerably lower feedback rate (around 60 percent, while the first round was around 80 percent).

What we can do, however, is to look at the quit rate. In case mini-jobbers would be considerably demotivated, one could expect an increase of their quit rate in the treatment shops, in particular, because they tend to have a much higher quit rate than regular employees. We do not find any evidence on this: the monthly quit rate of mini-jobbers is 5.1 percent in control versus 5.8 percent in the treatment, and the difference is statistically not significant (see Table 3, panel C; p-value: 0.448). The monthly quit rate of regular workers is 1.7 percent versus 2.1 percent (p-value: 0.250).

Leaving the interesting institutional specificity of the mini-jobbers behind, we would like to stress that team incentives may be a way to increase performance with limited effect on inequality.

VIII. Robustness

A. *Alternative Mechanisms*

Mechanisms that are unrelated to increased team effort cannot explain the treatment effect. *Working smarter* (Burgess et al. 2010) is no major driver of the treatment effect because reallocating labor hours between shifts takes time—at least a month under the company rules. The treatment effect, however, is stable in all months during the treatment period (see Table 7). *Extending opening hours* is impossible in 95 percent of the shops because they are located on the premises of large supermarkets. Rental agreements force them to exactly follow their host's opening hours, and it is physically impossible to remain open when the supermarket or mall they operate in is closing.[23] *Ordering more products* from the central warehouse to satisfy customer demand is impossible because the centralized ordering system gives little room for flexibility in orders. However, it reacts to an increase in sales by increasing the goods delivered in a proportional way. This explains why there is no treatment effect on the share of unsold goods (see Table 3, panel A).[24]

We also find no evidence that *shop supervisors' management input* is a major channel behind the treatment effect. The marginal increase in the shop supervisor bonus as a result of the team bonus—€80 per month on average—is rather small compared to the supervisors' individual bonus (up to €500 per month). Despite the sharp incentives already available to shop supervisors, sales targets were reached only in about 35 percent of the cases before the team bonus. Higher shop supervisor incentives with respect to sales, implemented in January 2014, did not produce a

[23] The only exception in the rental agreement is Sunday (German supermarkets have to be closed by law on Sundays). Removing from the analysis the shops that because of their architecture could be opened on Sundays does not change our results. Besides, assuming that the entire sales gain of 3 percent 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.

[24] We also find no treatment effect on mystery shopping scores (see Table 3, panel B), which does not come as a surprise. Historically, almost all shops reached in each month the maximum mystery shopping score of 100. In another project of ours, we find that mystery shopping scores had no predictive power for sales.

significant effect on sales.[25] Finally, the team bonus effect on sales is not affected by whatever proxy we use for shop supervisors' management input.[26]

B. *Contamination*

The heterogeneous treatment effects we find, in particular, in terms of team composition (share of the non-incentivized workforce) are hardly explainable by contamination. It is nonetheless important to note that we have taken great effort to prevent contamination between the treated and non-treated shops in our experimental design. This follows Bandiera, Barankay, and Rasul (2011), who argued that 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. We did not let the workers in the control group know that there was a team bonus in some other shops. Neither did the treatment group know there was a control group. We also developed communication protocols for the district managers to emphasize the fairness of the treatment assignment procedure. We learned from the district managers that only two employees from two control group shops asked them about the bonus. They received answers according to our protocol and seemed to be satisfied with the answer.

To detect potential contamination between treatment and control group shops we added questions about inter-shop employee contacts in the second wave of the employee survey. We found that there is not much inter-shop communication in general: 80 percent of the respondents almost never spoke to a colleague from another shop. Removing from our analysis the shops that had mentioned that they were communicating with employees from other shops did not affect the econometric results. We also studied 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. We could not find a single entry on the team bonus. As mentioned before, we also did not find any effect of the treatment in our employee survey, in which, among others, we asked questions about job satisfaction and organizational commitment.

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

[25] Before January 2014, shop supervisor bonuses depended on sales, personnel costs, and the mystery shopping score. Influenced by our findings that mystery shopping scores were too subjective to be a valuable performance indicator, management decided to remove the mystery shopping criterion from the rules determining the bonus (for all shop supervisors and before our treatment). 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 percent, and insignificant.

[26] These proxies are: supervisors' monthly working hours; tenure; average bonus received between January 2012 and March 2014; the leadership score (Carless, Wearing, and Mann 2000) from our employee survey; 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. While 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, 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 that neither workforce characteristics nor turnover changed in response to the bonus. Note also that these proxies for supervisor input are uncorrelated with the pretreatment share of mini-jobbers.

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 percent, close to the baseline, and the interaction coefficient is insignificant (p-value: 0.508). In summary, all of our contamination tests fail to provide any evidence of contamination.

C. *Gaming*

As mentioned previously, the step-wise bonus may lead to "gaming" (Courty and Marschke 1997), for example, through calibrating sales effort in order to just pass the bonus threshold. The risk of gaming is limited, because the incentive is relatively small (Gibbons 1998). However, we find a number of shops failing to reach their target by trivial amounts. For example, one shop failed to reach the target by €16, and another one by €8, which is an observation that is not consistent with gaming. We also 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 pretreatment sales volatility. Appendix B contains data and a more systematic discussion in support of gaming playing no role.

Another form of gaming could be present, because shop teams could consider working harder during the treatment period so to "convince" management to roll out the scheme. We do not believe this argument. First, as in Bloom et al.'s (2015) teleworking study, there are many small units in the treatment group. Because individual shops had little impact on the overall treatment effect, they had little incentive to exert effort beyond what their individual utility maximization required. Second, a number of pilot marketing initiatives had been introduced prior to 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.

Using the rollout episode, we can investigate whether the second type of gaming is likely to be present in the treatment shops. If gaming were the main reason behind the treatment effect, the treatment group's sales should decrease after the rollout to the control group level. Estimating 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), we find an effect of 4 percent which is comparable to our baseline treatment effect of 3 percent. Hence, it is likely that the control group increased their sales because of the rollout, rather than the treatment group decreasing theirs.[27]

[27] We also monitored that district managers did not spend more time with the treatment shops than with control shops. From the May 2014 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).

IX. Concluding Comments

Teams are a ubiquitous feature of modern production. Yet, so are monetary incentives that come in a multitude of shapes and sizes (Gibbons and Roberts 2012). MacLeod and Parent (1999) and Heywood and Parent (2012) report that around 15 percent of US companies paid bonuses to their workforce, and that between 10–20 percent of food service workers, and between 26–37 percent of sales workers received a bonus payment in their dataset, which, however, does not span recent data. For Germany, Heywood and Jirjahn (2002) find that 26 percent of companies use group incentives or profit sharing schemes. Despite a lack of detailed current data, it seems that bonus schemes in general, and in particular team bonus schemes are quite common in firms. Given the widespread discussion, team bonus schemes are likely to be on the rise, in particular in a world in which performance measurement is becoming easier.

The existing literature, discussed in the introduction, looked at team incentives in manufacturing and for teachers, and found evidence on efficiency-enhancing effects. In our team bonus experiment, we find positive efficiency effects on average, which benefit both the firm and the workers. The fact that workers also benefited is very important and in stark contrast to the evidence discussed in Hortaçsu and Syverson (2015) who established that the productivity gains in the US retail sector did not benefit workers. We believe our results to be applicable in a wide set of contexts; the kind of small teams we incentivized through our experiment are representative not only for retail but also many other service industries.

The heterogeneous treatment effects that we find suggest team incentives to work well where agents have sufficient leeway to influence the outcome variable (here, sales). Equally noteworthy is the result that those teams in which a sufficiently large part of total work hours is delivered by non-incentivized team members show no increase in performance at all.

Our experience makes us draw a number of conclusions on the questions asked by the literature on organizations and management practices. The foremost question is why not all firms in similar situations employ team incentives. Differences in product-market competition, knowledge, and organizational capabilities have been identified by the previous literature, and all of them certainly apply for our study. Potentially more interesting is that our study shows that it cannot be taken for granted that team incentives always work (as argued by Ichniowski and Shaw 2012). Initially, a number of the managers in our firm were concerned about the costs of the bonus, and they had good reasons for their skepticism. If the number of teams who increased their effort had been too small, the bonus would have been a loss-making activity: because the firm had committed itself to pay the bonus to all shops reaching the sales targets, this would have included those that would have reached the targets even in the absence of the bonus. It is in this context that running an experiment generates extra value: not only does the experiment allow to draw conclusions about the conditions under which efficiency gains are to be expected, but it also limits the stakes for a firm because it runs for a certain period, and only for a limited period of time (unless the firm decides to roll out the scheme, as was the case in our experiment). Together with a sense of crisis given the changing market structure, and with the support of the worker council, which strengthened the commitment power of management, this may have been the reason why we could overcome the initial concerns and organizational resistance.

Because of its simplicity, the bonus is relatively easy to implement. Its inequality effects were moderate. There is still much research to do, but employers should be made aware of the potential "win-win" benefits of providing team incentives in the workplace, and we would encourage a wider-spread use of RCTs in particular, in fields like retail in which there are many units that operate with the same technology, with a similar workforce, and under similar competitive situations, providing much statistical power and, at the same time, relevance for a substantial part of a country's workforce.

Appendix A: Proofs

PREDICTION VB: *A team's total effort decreases with the share of non-incentivized members in the team.*

Denote this share as 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 the individual one 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 nonlinear 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, Schivardi, and Tosetti 2008):

$$\text{(A1)} \quad E(e_1, \ldots, e_N) = \left(\sum_{i=1}^{N} e_i^{\rho}\right)^{\frac{1}{\rho}} \approx \left(\bar{e} + \frac{1}{2}(\rho - 1) \times \frac{\operatorname{var}(e)}{\bar{e}}\right) N^{1/\rho},$$

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 $\operatorname{var}(e) = (1 - \theta)\theta(e - 1)^2$ is the effort variance. Each incentivized worker's share in output is

$$\text{(A2)} \qquad \frac{\bar{e} + \frac{1}{2}(\rho - 1) \times \frac{\operatorname{var}(e)}{\bar{e}}}{1 - \theta} N^{\frac{1-\rho}{\rho}}.$$

Differentiating (A2) with respect to θ at $\theta = 0$ gives

$$\text{(A3)} \qquad \left(\frac{(\rho - 1) \times (e - 1)^2}{2e} + 1\right) N^{\frac{1-\rho}{\rho}}.$$

That is, under the incentivized workers' effort e being not too different from that or non-incentivized ones, and at ρ close to one, a small increase in θ from the base level of zero 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

$$\text{(A4)} \qquad \left(\frac{(\rho - 1) \times (e - 1)^2}{2e} + 1 - e\right) N^{\frac{1-\rho}{\rho}}.$$

Here, the positive effect on individual effort from the increase in individual shares in the total output (equation (A3)) is offset by the negative effect of replacing incentivized workers with less productive non-incentivized ones. The expression in (A4) 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 VC: *The effort under the bonus will increase with the distance between the actual and target sales before the bonus was introduced, provided this difference is not so large as to result in the corner solution* $e^* = e_0$.

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 $e_0 < e^* < e_{\max}$:

$$\text{(A5)} \qquad \left.\frac{de^*}{dy_0}\right|_{e=e_0} = -\frac{\dfrac{d^2\pi}{de_i\, dy_0}}{\dfrac{d^2\pi}{de_i^2}} = aN^{\frac{1-2\rho}{\rho}} B \frac{\dfrac{d\phi\left(aN^{\frac{1}{\rho}}e_0 - y_0\right)}{dy_0}}{\dfrac{d^2\pi}{de_i^2}}$$

$$= -aN^{\frac{1-2\rho}{\rho}} B \frac{\phi'\left(aN^{\frac{1}{\rho}}e_0 - y_0\right)}{\dfrac{d^2\pi}{de_i^2}}.$$

The derivative in (A5) 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 make 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. Whether the corner solution will occur depends on the costs of effort.

Appendix B

A. *Team Size and Response to Incentives in the Model*

Individual effort decreases with team size N if effort complementarities are not too strong $\left(\rho > \frac{1}{2}\right)$. 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,

$$\text{(A6)}\quad \left|\Phi''\left(aN^{\frac{1}{\rho}}e^* - y_0\right)\right| \ll \Phi'\left(aN^{\frac{1}{\rho}}e^* - y_0\right),$$

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

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

$$\text{(A7)}\quad \frac{d(Ne^*)}{dN} = e^* + N\frac{de^*}{dN}$$

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

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 (see also Esteban and Ray 2001 for the same result). Indeed, normalizing quantities to suppress the inessential parameters a, b, B, and y_0,

$$\text{(A8)}\quad \pi(e_i, e_{-i}) = \frac{1}{N}\left(e_i + \sum_{j\neq i} e_j\right) - e_i^2.$$

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

$$\text{(A9)}\quad \pi(e_i, e_{-i}) = \frac{1}{N}\left(e_i^\rho + \sum_{j\neq i} e_j^\rho\right)^{\frac{\gamma}{\rho}} - e_i^k, \quad k > 1,$$

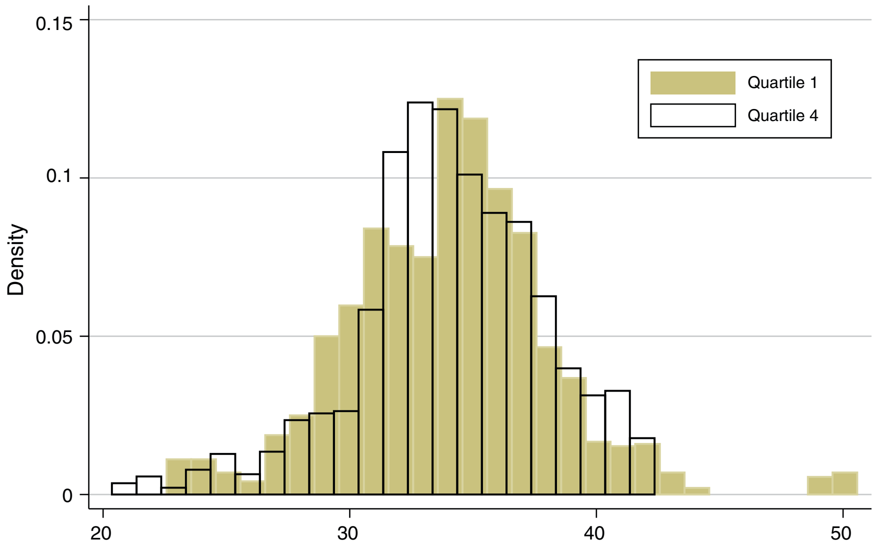

Figure A1. Distribution of Age in the First and Fourth Quartile of the Age Distribution

maximizing which with respect to effort results in the total effort equal to

$$(A10) \qquad N \cdot e^* = \left(\frac{\gamma}{k}\right)^{\frac{1}{k-\gamma}} \cdot N^{\frac{\gamma-2\rho}{\rho(k-\gamma)}+1}.$$

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

B. *Age Distribution in the First and Fourth Quartile*

Figure A1 plots the entire age distributions in the first and fourth quartile of average age in shops. We adjust by the differences in the mean age to aid comparison. The distributions differ: the one for Q1 has a longer right tail, which makes sense because the right tail in Q4 is naturally curtailed by people retiring. However, there are no other significant differences between the two age distributions.

C. *Gaming*

Figure A5 offers a systematic perspective on the symptoms of gaming by showing histograms of the log deviations of the actual sales from the target for the control and treatment groups separately. (For better visibility, only cases with the deviations within ± 10 percent are included.) As an indication for possible gaming, we observe 7.5 percent of cases with excess sales of between 0 and 0.5 percent in the treatment group and 4.5 percent 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

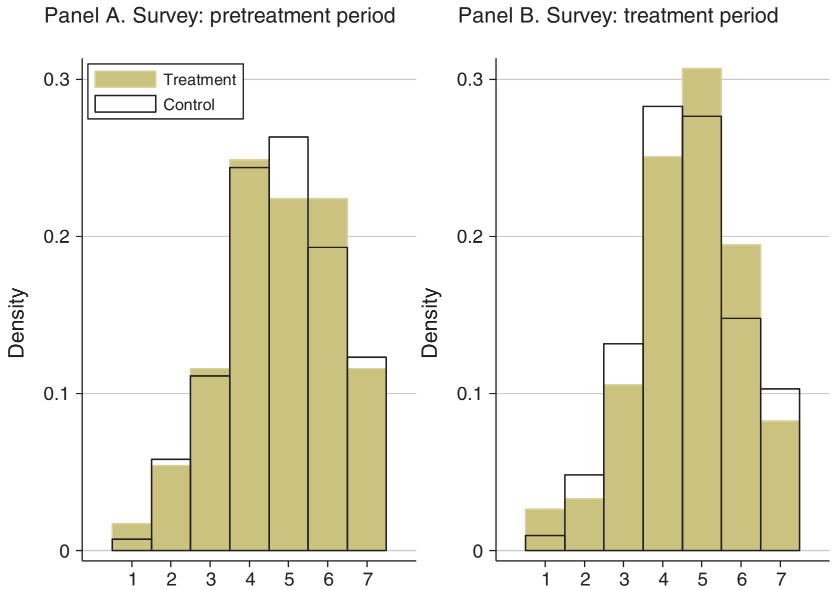

Figure A2. Distribution of Commitment Scores in Our Employee Surveys

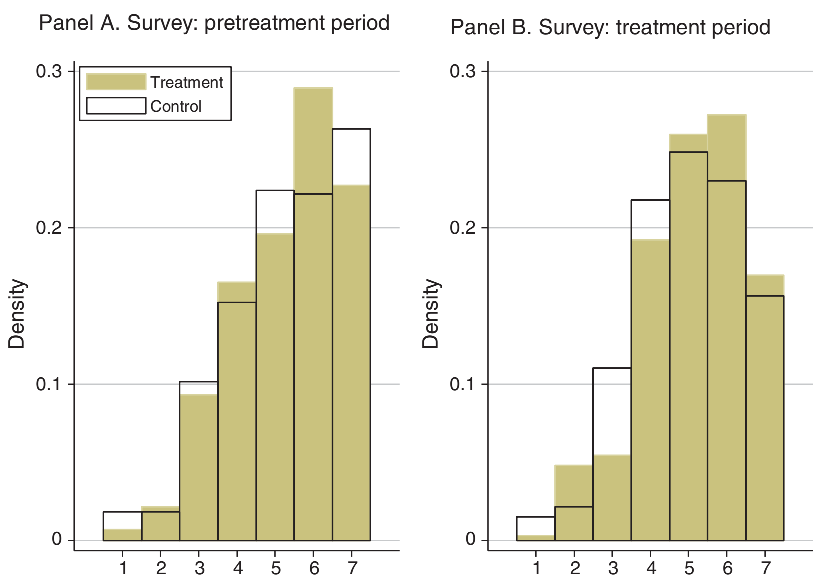

Figure A3. Distribution of Overall Satisfaction Scores in Our Employee Surveys

distributions in the treatment and control group once the treatment effect is subtracted from excess sales (p-value: 0.363). Second, there are no similarly prominent peaks at other cutoff points (1 percent, 2 percent, 3 percent, 4 percent 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

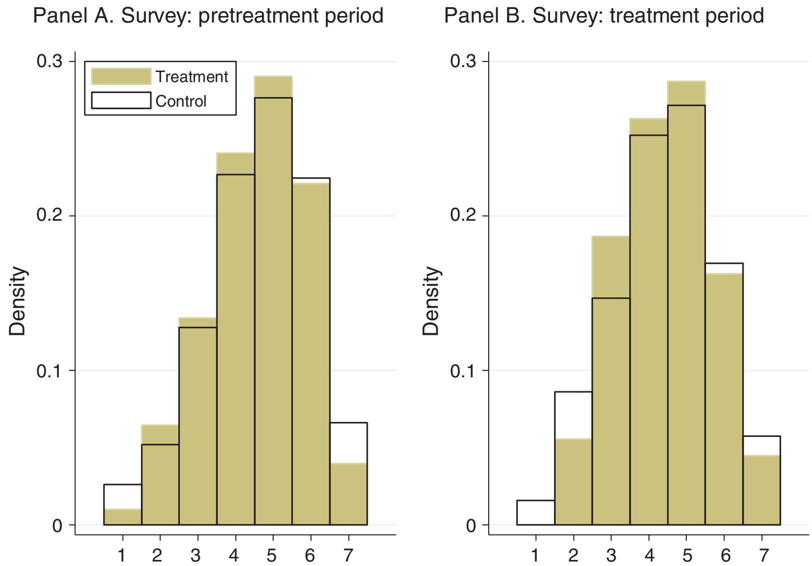

Figure A4. Distribution of Job Satisfaction Scores in Our Employee Surveys

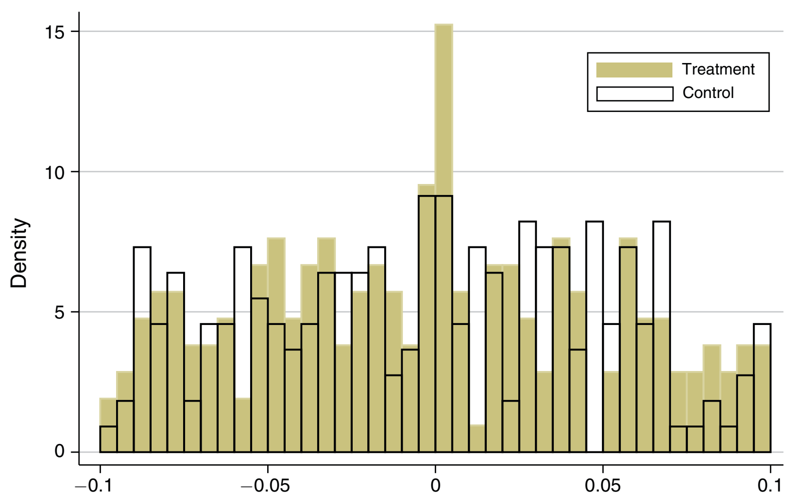

Figure A5. Percentage Deviation of Sales from the Target in April–June 2014

Notes: The graph plots histograms of percentage deviation of sales from the target for the treatment and control groups separately. For better visibility, only deviations within ±10 percent are included.

4.5 percent, a level at which no extra bonus is paid and gaming is unlikely (29.2 percent versus 23.6 percent in the treatment period). In fact, a naïve difference-in-difference calculation produces a borderline significant treatment effect of 0.076 on the frequency of excess sales above 4.5 percent. Summing up, the evidence for gaming is weak, and even if there is gaming, it would explain little of the treatment effect we have found.

REFERENCES

Albers, Sonke, Murali K. Mantrala, and Shrihari Sridhar. 2010. "Personal Selling Elasticities: A Meta-analysis." *Journal of Marketing Research* 47 (5): 840–53.

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62 (5): 777–95.

Allen, Natalie J., and John P. Meyer. 1990. "The Measurement and Antecedents of Affective, Continuance and Normative Commitment to the Organization." *Journal of Occupational and Organizational Psychology* 63 (1): 1–18.

Baily, Martin Neil, and Robert M. Solow. 2001. "International Productivity Comparisons Built from the Firm Level." *Journal of Economic Perspectives* 15 (3): 151–72.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2009. "Social Connections and Incentives in the Workplace: Evidence from Personnel Data." *Econometrica* 77 (4): 1047–94.

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

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2013. "Team Incentives: Evidence from a Firm Level Experiment." *Journal of the European Economic Association* 11 (5): 1079–114.

Barankay, Iwan. 2012. "Rank Incentives: Evidence from a Randomized Workplace Experiment." https://faculty.wharton.upenn.edu/wpcontent/uploads/2012/07/rankincentives_1.pdf (accessed June 6, 2017).

Barrios, Thomas. 2014. "Optimal Stratification in Randomized Experiments." http://scholar.harvard.edu/files/tbarrios/files/opstratv17_0.pdf (accessed June 6, 2017).

Bertrand, Marianne, and Francis Kramarz. 2002. "Does Entry Regulation Hinder Job Creation? Evidence from the French Retail Industry." *Quarterly Journal of Economics* 117 (4): 1369–413.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128 (1): 1–51.

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

Bloom, Nicholas, Renata Lemos, Rafaella Sadun, Daniela Scur, and John Van Reenen. 2014. "The New Empirical Economics of Management." *Journal of the European Economic Association* 12 (4): 835–76.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *Quarterly Journal of Economics* 130 (1): 165–218.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122 (4): 1351–408.

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

Bloom, Nicholas, and John Van Reenen. 2011. "Human Resource Management and Productivity." In *Handbook of Labor Economics*, Volume 4B, edited by Orley Ashenfelter, 1697–767. Amsterdam: Elsevier.

Boning, Brent, Casey Ichniowski, and Kathryn Shaw. 2007. "Opportunity Counts: Teams and the Effectiveness of Production Incentives." *Journal of Labor Economics* 25 (4): 613–50.

Burgess, Simon, Carol Propper, Marisa Ratto, Stephanie von Hinke, Kessler Scholder, and Emma Tominey. 2010. "Smarter Task Assignment or Greater Effort: The Impact of Incentives on Team Performance." *Economic Journal* 120 (547): 968–89.

Burks, Stephen, Jeffrey Carpenter, and Lorenz Goette. 2009. "Performance Pay and Worker Cooperation: Evidence from an Artefactual Field Experiment." *Journal of Economic Behavior and Organization* 70 (3): 458–69.

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

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

Cooper, Russell, and Andrew John. 1988. "Coordinating Coordination Failures in Keynesian Models." *Quarterly Journal of Economics* 103 (3): 441–63.

Courty, Pascal, and Gerald Marschke. 1997. "Measuring Government Performance: Lessons from a Federal Job-Training Program." *American Economic Review* 87 (2): 383–88.

Delfgaauw, Josse, Robert Dur, Arjan Non, and Willem Verbeke. 2014. "Dynamic Incentive Effects of Relative Performance Pay: A Field Experiment." *Labour Economics* 28: 1–13.

Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke. 2013. "Tournament Incentives in the Field: Gender Differences in the Workplace." *Journal of Labor Economics* 31 (2): 305–26.

Deloitte. 2016. *Global Human Capital Trends: The New Organizational: Different by Design*. London: Deloitte University Press.

de Zwart, Bart C. H., Monique H. W. Frings-Dresden, and Frank J. H. van Dijk. 1995. "Physical Workload and the Aging Worker: A Review of the Literature." *International Archives of Occupational and Environmental Health* 68 (1): 1–12.

Englmaier, Florian, Andreas Roider, and Uwe Sunde. Forthcoming. "The Role of Communication of Performance Schemes: Evidence from a Field Experiment." *Management Science*.

Esteban, Joan, and Debraj Ray. 2001. "Collective Action and the Group Size Paradox." *American Political Science Review* 95 (3): 663–72.

Friebel, Guido, Matthias Heinz, Miriam Krueger, and Nikolay Zubanov. 2017. "Team Incentives and Performance: Evidence from a Retail Chain: Dataset." *American Economic Review*. https://doi.org/10.1257/aer.20160788.

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

Friebel, Guido, and Levent Yilmaz. 2017. "Flexibility, Specialization and Individual Productivity: Evidence from Call Center Data." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2916042 (accessed June 6, 2017).

Frison, Lars, and Stuart J. Pocock. 1992. "Repeated Measures in Clinical Trials: Analysis Using Mean Summary Statistics and Its Implications for Design." *Statistics in Medicine* 11 (13): 1685–1704.

Gibbons, Robert. 1998. "Incentives in Organizations." *Journal of Economic Perspectives* 12 (4): 115–32.

Gibbons, Robert, and John Roberts. 2012. *The Handbook of Organizational Economics*. Princeton, NJ: Princeton University Press.

Hackman, J. R., and G. R. Oldham. 1980. *Work Redesign*. Reading: Addison-Wesley.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan. 2003. "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation." *Journal of Political Economy* 111 (3): 465–97.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4): 1009–55.

Heywood, John S., and Uwe Jirjahn. 2002. "Payment Schemes and Gender in Germany." *Industrial and Labor Relations Review* 56 (1): 44–64.

Heywood, John S., and Daniel Parent. 2012. "Performance Pay and the White-Black Wage Gap." *Journal of Labor Economics* 30 (2): 249–90.

Holmström, Bengt. 1982. "Moral Hazard in Teams." *Bell Journal of Economics* 13 (2): 324–40.

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

Hortaçsu, Ali, and Chad Syverson. 2015. "The Ongoing Evolution of US Retail: A Format Tug-of-War." *Journal of Economic Perspectives* 29 (4): 89–112.

Ichniowski, Casey, and Kathryn Shaw. 2012. "Insider Econometrics Empirical Studies of How Management Matters." In *Handbook of Organizational Economics*, edited by Robert Gibbons and John Roberts, 263–312. Princeton, NJ: Princeton University Press.

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

Iranzo, Susana, Fabiano Schivardi, and Elisa Tosetti. 2008. "Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data." *Journal of Labor Economics* 26 (2): 247–85.

Itoh, Hideshi. 1991. "Incentives to Help in Multi-Agent Situations." *Econometrica* 59 (3): 611–36.

Kandel, Eugene, and Edward P. Lazear. 1992. "Peer Pressure and Partnerships." *Journal of Political Economy* 100 (4): 801–17.

Kocher, Martin, Sabine Strauß, and Matthias Sutter. 2006. "Individual or Team Decision-Making-Causes and Consequences of Self-Selection." *Games and Economic Behavior* 2 (56): 259–70.

Laffont, Jean-Jacques, and Jean Tirole. 1988. "The Dynamics of Incentive Contracts." *Econometrica* 56 (5): 1153–75.

Lavy, Victor. 2002. "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." *Journal of Political Economy* 110 (6): 1286–317.

Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review* 90 (5): 1346–61.

Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent. 2009. "Performance Pay and Wage Inequality." *Quarterly Journal of Economics* 124 (1): 1–49.

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

List, John A., and Imran Rasul. 2011. "Field Experiments in Labor Economics." In *Handbook of Labor Economics*, Volume 4A, edited by Orley Ashenfelter and David Card, 103–228. Amsterdam: Elsevier.

MacLeod, W. Bentley, and Daniel Parent. 1999. "Job Characteristics, Wages, and the Employment Contract." *Federal Reserve Bank of St Louis Review* 81 (3): 13–27.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review* 99 (1): 112–45.

McKenzie, David. 2012. "Beyond Baseline and Follow-Up: The Case for More T in Experiments." *Journal of Development Economics* 99 (2): 210–21.

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

Nalbantian, Haig R., and Andrew Schotter. 1997. "Productivity under Group Incentives: An Experimental Study." *American Economic Review* 87 (3): 314–41.

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

Rademacher, Lena, Aallaa Salama, Gerhard Gründer, and Katja N. Spreckelmeyer. 2014. "Differential Patterns of Nucleus Accumbens Activation during Anticipation of Monetary and Social Reward in Young and Older Adults." *Social, Cognitive and Affective Neuroscience* 9 (6): 825–31.

Salthouse, Timothy A. 2010. "Influence of Age on Practice Effects in Longitudinal Neurocognitive Change." *Neuropsychology* 24 (5): 563–72.

Samanez-Larkin, Gregory R., and Brian Knutson. 2015. "Decision Making in the Ageing Brain: Changes in Affective and Motivational Circuits." *Nature Reviews Neuroscience* 16 (5): 278–89.

Shearer, Bruce. 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment." *Review of Economic Studies* 71 (2): 513–34.

Singh-Manoux, Archana, Mika Kivimaki, Maria M. Glymour, Alexis Elbaz, Claudine Berr, Klaus P. Ebmeier, Jane E. Ferrie, and Aline Dugravot. 2012. "Timing of Onset of Cognitive Decline: Results from Whitehall II Prospective Cohort Study." *British Medical Journal* 344.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49 (2): 326–65.

Tazhitdinova, Alisa. 2015. "Adjust Me If I Can't: The Effect of Firm Incentives on Labor Supply Responses to Taxes." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2648734 (accessed June 6, 2017).

Tellis, Gerard J. 2004. *Effective Advertising: Understanding When, How, and Why Advertising Works.* Thousand Oaks, CA: Sage Publications.

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

von Siemens, Ferdinand A., and Michael Kosfeld. 2014. "Team Production in Competitive Labor Markets with Adverse Selection." *European Economic Review* 68: 181–98.