Team Incentives, Social Cohesion, and Performance: A Natural Field Experiment

Josse Delfgaauw[1]
Robert Dur[1]
Okemena Abraham Onemu[2]
Joeri Sol[3]

[1] Erasmus University Rotterdam and Tinbergen Institute
[2] Leiden University
[3] University of Amsterdam and Tinbergen Institute

Team Incentives, Social Cohesion, and Performance: A Natural Field Experiment[*]

Josse Delfgaauw[†], Robert Dur[‡], Okemena Abraham Onemu[§], and Joeri Sol[¶]

June 2019

Abstract

We conduct a field experiment in a Dutch retail chain with 122 stores to study the interaction between team incentives, team social cohesion, and team performance. Theory predicts that the effect of team incentives on team performance depends on a team's social cohesion. In particular, free-riding should be weaker when co-workers care more about each other. Conversely, team incentives may lead to more co-worker support or to higher peer pressure and thereby can affect the team's social cohesion. We introduce short-term team incentives in a randomly selected subset of stores and measure for all stores, both before and after the intervention, the team's sales performance, the team's social cohesion as well as co-worker support and peer pressure. The average treatment effect of the team incentive on sales is 1.5 percentage points, which does not differ significantly from zero. In line with theory, the estimated treatment effect strongly increases in social cohesion as measured before the intervention. We find that social cohesion itself is not affected by the team incentives. Our study illustrates the potential of complementing a field experiment with ex ante and ex post questionnaire data collection for the study of management practices, workplace behavior, and performance.

JEL-codes: C93, M52.

Keywords: field experiment, team incentives, social cohesion.

[†]Erasmus University Rotterdam and Tinbergen Institute. E-mail: delfgaauw@ese.eur.nl.
[‡]Erasmus University Rotterdam, Tinbergen Institute, CESifo, and IZA. E-mail: dur@ese.eur.nl.
[§]Leiden University. E-mail: o.a.onemu@science.leidenuniv.nl.
[¶]University of Amsterdam and Tinbergen Institute. E-mail: j.sol@uva.nl.

1 Introduction

Teamwork is of vital importance in many professions, generating synergies in production, mutual learning, and social interactions with co-workers. But if co-workers are evaluated or rewarded based on team performance, free-riding may occur (Holmström 1982). Narrowly self-interested workers slack off as they ignore the benefits of their efforts that accrue to co-workers, yielding the well-known prediction that free-riding mutes the effect of team incentives on team performance. However, the classic free-rider argument ignores the role of social cohesion among co-workers, which might be a major deficiency given the high importance attached to it by both organizations and workers.[1]

In this paper, we combine field experimentation and collection of rich questionnaire data to study two key questions: How social cohesion among co-workers affects the effect of team incentives on team performance and, in turn, how team incentives affect social cohesion. Rotemberg (1994) and Dur and Sol (2010) extend the standard theory of free-riding in teams by including social cohesion, interpreted as co-worker altruism. The more employees care about their co-workers, the more they internalize the benefits of their efforts that accrue to co-workers. Hence, social cohesion reduces free-riding, implying that the effect of team incentives is larger in more cohesive teams.

We test this theoretical prediction by conducting a field experiment in a large retail chain in The Netherlands. We introduce team incentives in a randomly selected subset of stores. To filter out common shocks, the team incentive is designed as a contest between groups of 4 comparable stores. Within each group, all employees and the manager of the store that achieves the highest sales growth over a period of six weeks earn a monetary bonus. Just before the announcement of the team incentive, we conduct a survey among all employees of all stores, using a tried-and-tested survey scale to measure the social cohesion of employees within each store. This allows us to test whether the response to team incentives increases in the social cohesion of the team. In the analysis, we control for a large set of possibly confounding factors (i.e. store and team characteristics that may also affect the response to team incentives as well as correlate with social cohesion).

Shortly after the six-week incentive period, we conduct the survey again. This allows us to analyze how team incentives affect social cohesion in teams. In theory, this could go either way. Team incentives may induce team members to apply (more) peer pressure, to

---

[1]Organizations increasingly acknowledge the importance of social relations in the workplace. Many firms encourage social interactions between employees, for instance by organizing team bonding activities, adopting social technologies, and creating workplace designs to facilitate co-worker get-togethers (Waber et al. 2014, Deloitte 2016, McKinsey 2016, Gallup 2017). Employees value good social relations with co-workers. Social cohesion at work is a strong predictor of job satisfaction and organizational commitment, and is inversely related to absenteeism and employee turnover (Keller 1983, Mueller and Price 1990, Griffeth et al. 2000).

coerce their co-workers into exerting higher effort (Kandel and Lazear 1992, Barron and Gjerde 1997). Even if peer pressure is effective in raising team performance, social cohesion likely suffers. Alternatively, team incentives may induce team members to help each other at work (FitzRoy and Kraft 1986, Drago and Turnbull 1988) and to invest (more) in co-worker social relations (Rotemberg 1994, Dur and Sol 2010). This would raise both performance and social cohesion. To analyze whether these mechanisms are active, we include measures of peer pressure and co-worker helping in our surveys.

The literature on cohesion dates back to Festinger (1950), who defines group cohesion as "the resultant of all the forces acting on members to remain in the group." (Festinger 1950, p. 274). Subsequent literature distinguishes between task cohesion and social cohesion (Zaccaro and Lowe 1988; Mullen and Copper 1994; Casey-Campbell and Martens 2009, Chiocchio and Essiembre 2009).[2] Task cohesion refers to group members' joint commitment to the group's tasks or goals, whereas social cohesion captures interpersonal attraction to the group. This includes friendship relations and the extent to which group members enjoy spending time together. Following Widmeyer et al. (1995), Carless and De Paola (2000, p.73) define social cohesion as "the motivation to develop and maintain social relationships within the group." To measure cohesion, we use the survey scale developed by Carless and De Paola (2000), discussed in more detail in Section 2.3.[3]

Our key findings are as follows. First, the introduction of team incentives increases sales by 1.5 percentage points. This average treatment effect is not statistically significant, although it falls well within the range of findings in related studies in similar contexts. As predicted, the effect of the team incentive on sales increases in the team's pre-existing level of social cohesion. This effect is sizeable: a one standard deviation increase in social cohesion increases the estimated effect of the team incentives on sales by 2 percentage points. We find a statistically significant effect of team incentives on sales for stores that belong to the top quartile in terms of social cohesion. Hence, we provide strong evidence that better social relations among team members mitigate free-rider effects.

Using the pre- and post-experiment questionnaire data, we find a small and statistically insignificant positive effect of team incentives on social cohesion. We do find limited evidence

---

[2] A limited number of studies also include group pride as a component of group cohesion (Mullen and Copper 1994; Chiocchio and Essiembre 2009).

[3] What drives differences in social cohesion between work teams? Correlational studies document that groups that are more homogeneous in terms of personal characteristics, such as gender, ethnicity, age, and tenure, tend to be more cohesive (Williams and O'Reilly 1998; McPherson et al. 2001). Cohesiveness relates to leadership (Nishii and Mayer 2009; Wendt et al. 2009), to group size (Carron and Sprink 1995), and to the type of tasks the team performs (Zaccaro and McCoy 1988). We control for most of these possible confounds in our empirical analysis. Relatedly, Harrison et al. (2002) and Berger et al. (2011) find that team members collaborate and cooperate more in the presence of team-based performance pay. Employees are also more likely to engage in peer pressure, i.e. to take action upon observing a shirking colleague, in firms that offer team incentives (Freeman et al. 2010).

for an increase in co-worker helping, and find no effects on peer pressure and job satisfaction. Hence, social relations among team members are neither improved nor hurt by the implementation of team incentives.

Our paper contributes to two strands of literature. First, we contribute to a rapidly expanding literature studying the effects of incentive pay in the workplace (for recent reviews see Levitt and Neckermann 2014 and Lazear 2018). In an early field experiment, Erev et al. (1993) documents substantial free-riding under team incentives among orange pickers. Erev et al. (1993) and Bandiera et al. (2013) find that a contest between teams leads to higher performance than team incentives based on the team's absolute performance. This is also a common finding in lab experiments, see Sheremeta (2018). Lavy (2002) finds positive effects of team-based pay for teachers on student performance, but subsequent studies have found mixed results (Glewwe et al. 2010; Muralidharan and Sundararaman 2011, Fryer 2013). Team incentives increased performance at the UK tax authorities, partially through the reallocation of more able employees to incentivized tasks (Burgess et al. 2010). Englmaier et al. (2018) finds positive effects of team incentives on group performance in escape rooms.

Several recent studies analyze the effects of team incentives in retail chains. Using comparable levels of incentive pay, average treatment effects on performance range from 0 to 5 percent (Delfgaauw et al. 2013, 2014, 2015, 2019; Friebel et al. 2017). Among others, these papers show that the response to team incentives relates to the gender composition of the team (Delfgaauw et al. 2013) and to measures of local demand and to the share of non-eligible employees (Friebel et al. 2017). None of these studies analyzes the interplay between team incentives, social cohesion, and performance.

Second, we add to the literature on social incentives in the workplace (Ashraf and Bandiera 2018), which studies how social concerns affect workers' performance. A classic study in this area is Hamilton et al. (2003), which finds increased output after a switch from individual production and incentives to team production and incentives in a garment factory. Some employees voluntarily joined teams despite a drop in earnings. Bandiera et al. (2005) finds that switching from relative performance pay to individual performance pay increases performance among fruit pickers. This effect is most pronounced among socially related workers, suggesting that workers (partially) internalize the negative externality they impose on others.[4] Comparing individual and team incentives for gym and library visits, Babcock et al. (2015) finds higher performance for teams when team members know each other. Our results are in line with the insight that individuals in teams with good social

---

[4]Incentives can also affect (self-)selection. Bandiera et al. (2009) finds that managerial incentive pay induces managers to select high-ability workers rather than friends. Studying endogenous team formation, Bandiera et al. (2013) find that team incentives induce workers to sort into teams on the basis of ability rather than friendship. In our firm, teams exist before the introduction of team incentives, implying that (self-)selection is of minor importance.

relations internalize external effects on team members. Furthermore, we are one of the first to analyze how team incentives affect social relations. The only other study we are aware of that has a similar objective is Carpenter and Seki (2011), studying fishermen in Toyoma Bay. Part of their findings are close to ours: fishermen who engage in team production are more productive than those who do not and this can be largely attributed to social preferences. In contrast to our results, they do find an effect of team production on team members' altruism. Given that many firms and employees value social relations at work (see footnote 1 above), it is surprising how little evidence exists on how organizational policies affect social relations among employees.

Given the absence of exogenous variation in social cohesion in our study, we cannot assess whether strengthening social cohesion leads to better team performance. Meta-analyses of the literature in organizational psychology distill a positive relation, but acknowledge mixed results of individual studies (Mullen and Copper 1994, Gully et al. 1995, Beal et al. 2003, Bell 2007, Chiocchio and Essiembre 2009). Chen and Lim (2013) and Lim and Chen (2014) show that allowing participants to interact before performing a task in a lab experiment improves performance under team incentives but not under individual incentives. Corgnet et al. (2019) finds that under team incentives, but not under individual incentives, participants perform worse if one of their co-workers is a robot rather than a human. This pattern is stronger for altruistic than for non-altruistic individuals, highlighting the interaction between social and material incentives. De Paola et al. (2019) conduct a field experiment and find that students assigned to teams composed of friends perform better on both individual and team assignments than students assigned to team composed of strangers.

Our study also relates to the literature on group identity and in-group favoritism (Tajfel and Turner 1979, Akerlof and Kranton 2000). Experimental studies find that people behave more altruistically and cooperatively towards members from the same group than to outsiders, even if groups are randomly formed (Goette et al. 2006, Charness et al. 2007, Chen and Li 2009). In situations of intergroup conflict, group members may even be hostile towards members of other groups (see e.g. Goette et al. 2012). We study an actual workplace setting, and show that employees' perception of their group's cohesiveness matters for how they respond to a monetary incentive with both positive in-group externalities and negative out-group externalities.

For managers, our findings indicate complementarities between team incentives and social cohesion in production. Hence, investments in social cohesion yield higher returns if team incentives are stronger. Furthermore, the team incentives neither improved nor harmed social cohesion within stores. This may comfort managers who consider introducing team incentives but worry about the effects on social relations within the team. At the same time,

our results suggest that team incentives are no panacea for bad co-worker relations.

Methodologically, our paper illustrates the potential of a research design that combines field experimentation and rich questionnaire data collection at multiple points in time. It allows to estimate heterogeneous treatment effects of policies with respect to individual or team characteristics that are best measured using questionnaires, and to estimate the effects of specific policies and practices on those characteristics. This holds promise for deepening our understanding of the mechanisms that drive the effects of management practices on organizational performance, and complements the approach of measuring and improving the general quality of firms' management practices (Bloom and van Reenen 2007; Bloom et al. 2013).

The remainder of the paper proceeds as follows: Section 2 presents our experimental setting and design, Section 3 describes our data and Section 4 contains our methodology. We present the main results in Section 5 and discuss several robustness checks in Section 6. Section 7 concludes with a discussion.

2 The Experiment

2.1 The firm

We conduct the field experiment in a retail chain consisting of 122 stores during the Fall of 2013.[5] All stores of the retail chain are located in The Netherlands and operate under a single brand name. The stores sell clothes, shoes, and items for sports and outdoor activities. The chain mostly targets budget-minded consumers. In the company, many decisions are taken by the central management at the headquarters. The product range, pricing, general personnel policies, the internal store design, branding, and advertisement are uniform across stores. Store managers are responsible for day-to-day organization and for the staffing of their store. Within a store, products are placed on displays. Customers can take these items to the registry. Alternatively, customers can ask an employee for help, for instance to measure their shoe size or to see whether a particular item is available.

Stores vary in size, and employ a manager and between 5 and 17 employees.[6] As many employees work part-time or on-call, on a typical weekday only a subset of employees are present in the store. Employees earn a flat wage just above the legal minimum wage. Store managers earn about 40% more, and a small fraction of their pay is performance-related. Employees' tasks include manning the registry, keeping the displays stocked and tidy, advis-

---

[5] The results of two earlier experiments within the same retail chain are reported in Delfgaauw et al. (2011, 2013). These field experiments took place between 2007 and 2009.

[6] During our experiment, two stores did not have a store manager.

ing customers, and cleaning the store. In this environment, employees' actions and efforts will not have enormous effects on sales. As mentioned in the Introduction, earlier experiments in similar settings find that incentive pay for employees tend to increase sales by about 0 to 5 percentage points. The chain's central management is confident that employees can affect sales, for instance by keeping displays stocked, providing good service and advice to customers, and by suggesting alternative or complementary products. For this reason, they occasionally organize short-term incentive events.

2.2 Experimental design

The experimental treatment is the introduction of a monetary team incentive for the employees of 72 randomly chosen stores. The remaining 50 stores make up our control group. We decided to make the treatment group larger than the control group to increase power, anticipating that the variance of performance during the treatment period would be larger among stores in the treatment group than among stores in the control group (List et al. 2011). To achieve balance of the key variables for our analysis, we stratified on both stores' social cohesion (as measured in the pre-experiment survey) and stores' past sales peformance. We explain the randomization procedure in subsection 2.4 below.

The 72 stores assigned to the treatment group are subdivided into groups of 4 stores. Within each group, all employees and the manager of the store that achieves the highest sales growth over a period of six weeks receive a monetary bonus. Sales growth is measured as the percentage change in sales compared to the same weeks a year earlier. This is a common performance measure in this retail chain. Employees are generally well-aware of their store's performance in terms of sales growth. The bonus was set at 75 euro for full-time employees, 50 euro for part-time employees, and 25 euro for on-call employees.[7] For full-time employees, the bonus is about 4% of their monthly earnings, corresponding to about 2.8% of earnings over a six-week period. For store managers, the bonus is about 2.9% of monthly earnings. These amounts are comparable to bonus pay in earlier experiments in retail chains (see the Discussion section in Delfgaauw et al. 2019). The treatment period was week 42 to 47 in 2013 (October 14th to November 24th). During this period, no other events or policy changes took place. Hence, for stores in the control group, it was business-as-usual.

We deliberately designed the team incentive as a contest between stores. Stores' sales are highly volatile, largely due to chain-wide factors such as holidays, weather, and advertisement campaigns. Figure 1 shows the weekly average of all stores' sales growth for the period of

---

[7]Employees would receive this bonus independent of whether they were actually at work, implying that (self-)selection effects are likely limited.

Figure 1: Weekly sales growth
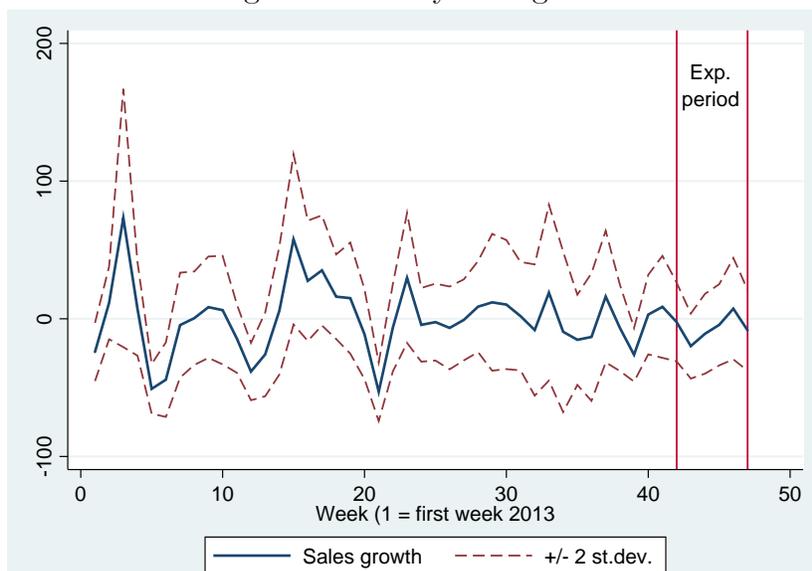
week 1 to week 47 in 2013.[8] Week-fixed effects alone explain 58.4% of the total variation in stores' sales growth over this period, underlining the importance of common shocks to performance. The large volatility in sales makes it undesirable to use absolute sales targets. A positive (negative) shock can make a pre-determined target too easy (difficult) to reach, reducing the incentive effect. As a large part of volatility is due to common shocks in our context, relative performance pay is more suitable for providing incentives because its incentive effect is immune to common shocks (Lazear and Rosen 1981, Nalebuff and Stiglitz 1982, Green and Stokey 1983).

The effects of tournament incentives are predicted to be stronger if contestants are more homogenous in ability (Lazear and Rosen 1981). Therefore, we grouped stores on the basis of historical performance, as follows. We ranked the stores on the basis of sales growth in the six months preceding the treatment period (week 15 to 40, 2013). The top four stores comprised one tournament group, as well as stores ranked 5th to 8th, and so on. To limit the scope for collusion or sabotage, we adjusted the group composition if stores from the same region were grouped together.[9]

All communication on the experiment to employees ran through the company's regular communication channels. Two weeks before the start of the contests, a general announcement was sent digitally to all employees. The message stated that an incentive event would take place in a few weeks, that a subset of stores would participate in this event, and that more information would follow soon. Three days before the start, all store managers received

---

[8]We do not have access to sales growth data after the experimental period.

[9]We made 7 such adjustments.

Figure 2: Example of the poster stores received every week (translated)

Ranking after 3 weeks:

1. **Amsterdam**   **(8121)**   **- 3.7 %**

2. **Rotterdam**   **(8024)**   **- 6.9 %**

3. **Hengelo**   **(8030)**   **- 7.5 %**

4. **Best**   **(8103)**   **- 14.6 %**

Percentages are realized sales growth compared to the same period in 2012.
The contest lasts 6 weeks. Fulltime employees of the winning store receive 75 euro per person! Parttime employees of the winning store receive 50 euro per person! On-call employees of the winning store receive 25 euro per person!

a (hardcopy) letter, signed by the sales manager and the HR manager of the retail chain. The letter for managers of stores in the control group stated that their store would not participate this time, but would participate in a later event.[10] The letter for managers of stores in the treatment group stated that their store would participate in the current contest, and explained the rules and the reward.

Furthermore, on the first day of the contest, all participating stores received a large, brightly-colored poster designed for this event. The poster stated the period of the sales contest, listed all four stores in the group (in alphabetical order), and mentioned that the group composed of stores with similar sales growth in the past period. During the contest, all participating stores received weekly feedback, in the form of a large poster containing a ranking of the stores and their sales growth in the contest so far. These posters also reminded employees of the reward for winning. See Figure 2 for an example of these posters. Store managers were instructed to hang the posters on a prominent place in the store's canteen.

Stores in the control group did not receive any posters. This implies that our intervention comprises the team incentive and the weekly posters.[11] As sales growth is a well-known

---

[10]Employees and managers of different stores do not regularly communicate, but do occasionally interact. Hence, if stores in the control group would not have been informed ex ante about the incentive event, it is likely that some of them would have learned about it during the experiment. To prevent stores' response to such surprises from affecting our estimations, we decided to inform the control group about the contest. As the retail chain conducts incentive events more often, and sometimes for subsets of stores, we do not expect that informing the control group leads to a contamination bias.

[11]Englmaier et al. (2016) shows that (better) communication about the incentive scheme can dramatically increase its effect on performance.

performance measure in the retail chain, we have little reason to believe that any treatment effects are also driven by better access to information on performance.

2.3   Survey design

A key variable for our analysis is the social cohesion of employees within a store. We are interested in how pre-existing social cohesion affects the response to team incentives, as well as whether and how social cohesion is affected by the introduction of team incentives. This requires measuring social cohesion, both before and after the experiment. Furthermore, we are interested in whether the effect of team incentives on social cohesion relates to effects on co-worker helping and/or on peer pressure. For this reason, we administer a pre-experiment and a post-experiment survey among employees and store managers. For all survey measures, we used tried-and-tested survey scales, as discussed below. The complete survey can be found in the Appendix.

Several survey measures for teams' social cohesion exist in the literature on group cohesion following Festinger (1950). We use the survey questions developed by Carless and De Paola (2000). They adapted the group cohesion scale by Widmeyer et al. (1985) from a sports setting to a work setting.[12] Table 1 lists the six survey items that together form the measure of social cohesion (our full survey can be found in the Appendix). Each item is a statement to which employees can respond on a 7-point Likert scale ranging from "completely disagree" to "completely agree". The questions capture whether work group members like to interact outside of work and whether this team is an important social group for the respondent. We take the average of the responses to the six questions as an individual's perception of his or her team's social cohesion.[13]

Social cohesion has been shown to correlate strongly with task cohesion and leadership style (Mullen and Copper 1994, Carless and De Paola 2000, Wendt et al. 2009, Chiocchio and Essiembre 2009). We included measures of task cohesion and leadership style in our survey so as to be able to control for treatment effect heterogeneity with respect to these variables in our analysis. Task cohesion captures the degree of alignment among team members regarding the team's goals and approach, and is generally measured alongside social cohesion in determining group cohesion. We implement the survey scale developed by Carless and De Paola (2000), which contains 4 items. Wendt et al. (2009) show that directive (supportive) leadership is negatively (positively) correlated with group cohesion. Directive leadership captures a management style of control, close supervision, and authority. Sup-

---

[12]Price and Mueller (1986) offer an alternative social cohesion scale. See Casey-Campbell and Martens (2009) for a discussion on measurement of group cohesion.

[13]We reverse the responses of the negatively-phrased questions, so that a higher response corresponds to higher social cohesion. The same holds for the other survey measures.

Table 1: Survey questions on social cohesion, adapted from Carless and De Paola (2000)

Please indicate for each proposition which answer corresponds best to your situation in the past two months.

The following questions revolve around the atmosphere in your store.
- Our team would like to spend time together outside of work hours.
- Members of our team do not stick together outside of work hours.
- Our team members rarely do something nice together.
- Members of our team would rather go out on their own than get together as a team.
- For me this team is one of the most important social groups to which I belong.
- Some of my best friends are in this team.

The propositions came with answers on a 7-point Likert scale ranging from 'Completely agree' to 'Compeletely disagree'.

portive leadership revolves around building relationships, encouragement, and concern for employees. To reduce the length of the survey, we use 6 out of the 14 questions in the survey scale used by Wendt et al. (2009), 3 for both leadership styles.

In the literature on group cohesion, there is discussion whether cohesion is an individual-level or a group-level construct (Dion 2000, Friedkin 2004). As our main intervention is a team incentive and, hence, administered at the team level, we choose to create measures of social cohesion, task cohesion, and leadership style at the team-level as well. Thereto, we take the mean of the individuals' average score on a given survey measure over all respondents from a given store. We use these store-level measures to determine how the effect of team incentives depends on social cohesion in the store. In the analysis of the effect of team incentives on social cohesion, we estimate the effect both using individual-level data and using store-level data.

For peer pressure, we use the survey questions developed by Freeman et al. (2008, 2010). Their 'anti-shirking index' measures the willingness to undertake action upon observing a shirking colleague, where employees could indicate the likelihood that they would (a) do nothing, (b) talk to the shirking colleague, (c) talk to other colleagues, and (d) talk to the supervisor.[14] Using cross-sectional data from the General Social Survey in the US, Freeman et al. (2010) finds that a majority of workers state that they would take action upon observing a shirking colleague. Furthermore, they find that anti-shirking behavior is positively related to the presence of team incentives.

For co-worker helping, we use 2 out of 5 questions from the scale for interpersonal helping developed Moorman and Blakely (1995), who used this scale as one dimension in their measurement of organizational citizenship behavior. The questions capture the extent of

---

[14]Before asking the questions on peer pressure, we ask whether it is easy for employees to determine whether their colleagues work hard. Mean response is 5.3 on a 7-point scale, which indicates that most employees think they can observe whether someone shirks or not.

team members' willingness to help each other at work. Lastly, we include three questions that measure individuals' job satisfaction, job search, and intention to quit. For job satisfaction, we use a general question as in Clark (2001). For job search and intention to quit, we follow the recommendation by Tett and Meyer (1993) to use single-item measures and to refer to a specific time period. The question for job search is retrospective: "Did you search for another job in the past two months?" Here, we chose the two-month period as this corresponds to the period between the two surveys (in which the experiment took place). The question for intentions to quit is forward-looking instead. Survey questions on job search and intention to quit have been shown to predict actual quits (Böckerman and Ilmakunnas 2009, Cornelissen 2009, Card et al. 2012).

Employees and store managers receive almost identical surveys. There are two differences. First, in the measure for peer pressure, the survey for store managers does not contain the question on whether the respondent reports a shirking colleague to the store manager. Second, in the questions on leadership, 'The store manager' is replaced with 'I'. In the main analysis, we make no distinctions between managers and employees in determining team's social cohesion and other store-level survey outcomes. As a robustness check, we also construct store-level measures excluding the manager.

We send e-mail invitations to take part in the pre- and post-experiment online questionnaires to all employees and store managers. This e-mail stated that the company collaborated with the University of Amsterdam to investigate the satisfaction of all employees in the stores. It also stated that the survey would take less than 5 minutes to complete, and that responses would not be shared with the company, neither at the individual nor at the store level. We offered an incentive to complete both surveys: we allotted a tablet (retail price 200 euro) to an individual randomly drawn from the set of people who completed both surveys. For each survey, employees were given 10 days to complete it. We sent two reminders to non-respondents. For 16 percent of the employees, we did not have an email-address. These employees received the invitations through regular mail. The invitation letters were identical to the e-mails, apart from a personalized password to access the questionnaire online. We did not send reminders through regular mail.[15]

---

[15]It is unlikely that employees linked the questionnaires to our experiment, as the company has organized similar sales competitions in the past. Moreover, neither of the questionnaires nor the accompanying emails mentioned the sales competitions or team incentives. Hence, we maintain that our experiment can be classified as a natural field experiment according to the taxonomy of field experiments by Harrison and List (2004).

Figure 3: Design and timeline of the experiment

Bullet (square) indicates a point of random (non-random) assignment.
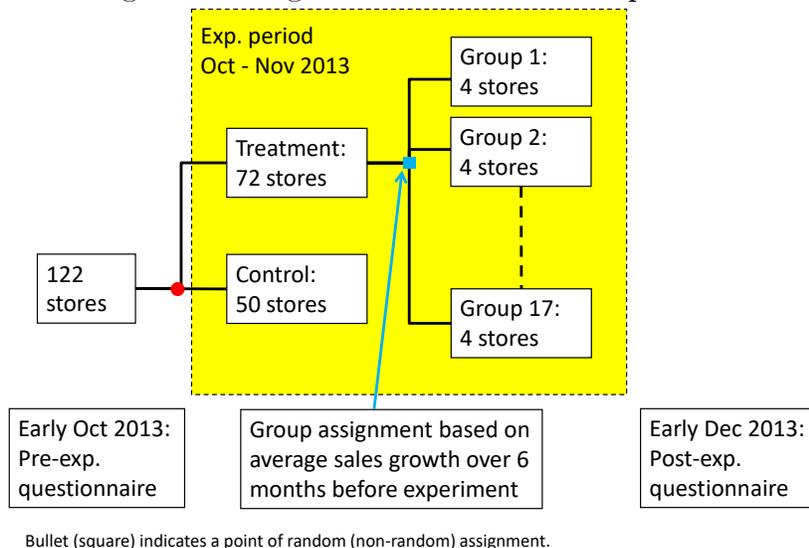
2.4 The assignment procedure

In designing our randomization procedure, we aimed to make the treatment group and the control group similar in their distributions of sales performance and social cohesion. Thereto, we stratified on social cohesion as measured in the pre-experiment survey and on past performance, as follows. First, we ranked all stores on their social cohesion.[16] We created 12 blocks of stores with similar social cohesion, i.e. 10 blocks with 10 stores and 2 blocks with 11 stores. Next, we ranked all stores within each block on the basis of average sales growth over the preceding six months (weeks 15 - 40, 2013). We split each block into two equally large strata, where one stratum contained the stores with relatively high sales growth and the other stratum the stores with relatively low sales growth. Finally, we randomly assigned three stores from each of the 24 strata to the treatment group, yielding 72 treatment stores. The remaining 50 stores make up our control group.

Figure 3 presents a schematic overview of the design and the timing of events.

3 Descriptive Statistics

Our analysis draws on three sources of data. First, we use weekly data on sales for each store for the period of week 1 to week 47 in 2013. For reasons of confidentiality, we do not have access to absolute sales figures. Instead, we use indexed sales data for the period of

---

[16]For two stores, we did not obtain any response on the pre-experiment questionnaire, implying that these stores did not have a social cohesion score. For both stores, we drew a random score from the range of scores of the other stores, and used this to place the stores in the ranking. We discuss survey non-response in more detail in Section 3.2.

week 1 in 2012 to week 47 in 2013. Week 1 in 2010 is the base week for each store. For our analysis, this allows us to compute a store's sales growth compared to the same week a year earlier.[17] Note that this performance measure is also used in the sales competition. Second, we use the companies' personnel data as of September 2013, containing demographic and contractual information for all employees. This includes employees' gender, age, tenure, and job level. Third, we use the data obtained from the pre- and post-experiment surveys conducted among the employees.

3.1   Store and personnel characteristics

The first column in Table 2 reports descriptive statistics regarding store performance and personnel characteristics. Average sales growth before the experimental period is close to zero, but differs substantially across stores. During the experimental period, average sales growth is negative, which is not extraordinary as can be seen in Figure 1. Like many other retail chains, ours suffered from the aftermath of the economic crisis as well as from increased competition from online shops. Within-store standard deviation of sales growth across weeks is high, echoing the volatility of sales growth depicted in Figure 1. On average, stores have about 10 employees, and more than five out of six employees are female. Employees are relatively young and have on average 6 years of tenure. About 40 percent of employees works on-call. The average manager is older and has a longer tenure than the average employee. Almost 60 percent of store managers is male.

Columns 2 and 3 in Table 2 show that stores in the treatment group and stores in the control group are similar in terms of sales growth before the experiment as well as on all store characteristics. Average sales growth before the experiment is slightly, but not significantly, higher in the treatment group than in the control group. This difference increases somewhat during the experimental period, heralding the average treatment effect we report in our analysis below.

3.2   Survey data

3.2.1   Descriptives

Panel A and panel B of Table 3 give the results of the pre- and post-experiment survey, respectively. Column 1 shows that on average, 35 percent of the employees in a store completed

---

[17]Let $R_{s,w,y}$ be sales of store s in week w of year y, and $R_{s,b}$ the sales of store s in the base week. We receive index $I_{s,w,y} = \frac{R_{s,w,y}}{R_{s,b}}$. Hence, sales growth of store s in week w in 2013 relative to the same week a year ealier is calculated as $g_{s,w} = \frac{I_{s,w,2013} - I_{s,w,2012}}{I_{s,w,2012}} = \frac{\frac{R_{s,w,2013}}{R_{s,b}} - \frac{R_{s,w,2012}}{R_{s,b}}}{\frac{R_{s,w,2012}}{R_{s,b}}} = \frac{R_{s,w,2013} - R_{s,w,2012}}{R_{s,w,2012}}$, which is independent of base week sales $R_{s,b}$.

Figure 4: Distribution of social cohesion across stores
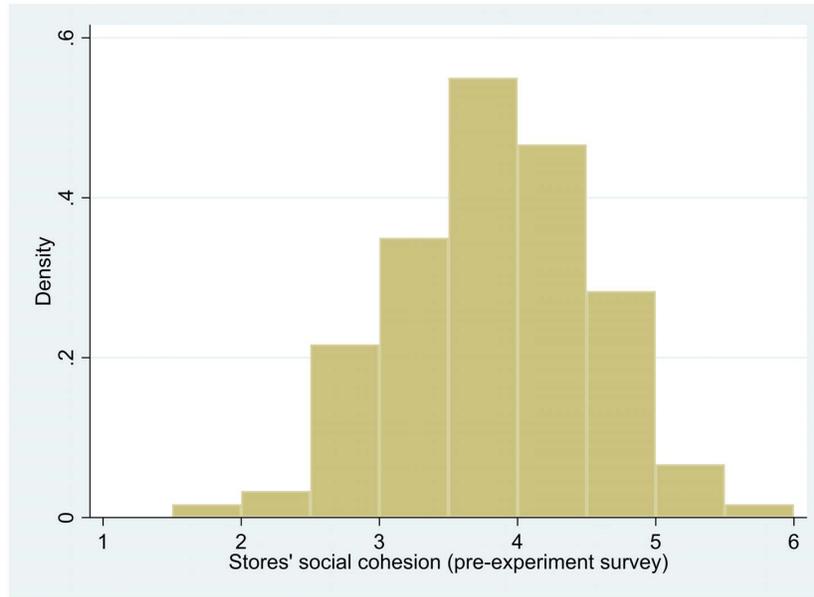
the first survey, as well as about four out of five managers. The second survey is completed by 23 percent of employees and 68 percent of managers.[18] We discuss self-selection into the surveys below.

The average response to the survey questions on social cohesion in the pre-experiment survey is 3.8 on a 7-point scale. Figure 4 gives the distribution of stores' social cohesion, showing considerable variation. Hence, in some stores, the team is very close and engages in social activities outside of work hours, whereas in other stores, people do not regard their colleagues as an important social group.

The other survey measures in the first column in Panel A of Table 3 show that task cohesion is stronger than social cohesion, that leadership is perceived as more supportive than directive, that employees experience substantial co-worker helping, and that employees impose a moderate level of peer pressure. Columns 2 and 3 in Panel A show that response rates and all survey measures are very similar across the treatment group and the control group. Hence, our assignment procedure has created two groups of stores that are, on average, similar in historical performance, personnel characteristics, and all survey measures. Panel B of Table 3 shows that in the post-experiment survey, the outcomes are quite close to the pre-experiment outcomes.[19]

Table 4 provides correlations between store performance, personnel characteristics, and

---

[18]Table 3 gives the average of outcomes at the store level. The overall response rate of employees to the pre- and post-experiment surveys is also 35 percent and 23 percent, respectively.

[19]In the data, job satisfaction, job search and intention to quit are strongly correlated. For brevity, we have chosen to only report job satisfaction.

the survey measures (averaged at the store level). Average sales growth before the experiment is neither correlated with social cohesion nor with task cohesion, but is positively (negatively) correlated with supportive (directive) leadership.[20] None of the personnel characteristics correlates strongly with store's sales growth. In line with the earlier literature, social cohesion correlates positively with both task cohesion and supportive leadership and negatively with directive leadership. As was to be expected, co-worker helping correlates positively with social cohesion, whereas peer pressure correlates negatively. Personnel characteristics do not correlate with social cohesion, except for a negative correlation with the dispersion in age among employees. More generally, store characteristics do not correlate strongly with the survey measures. In line with earlier studies (Williams and O'Reilly 1998; McPherson et al. 2001), we do find negative correlations between task cohesion and the dispersion in tenure and age among employees as well as between co-worker helping and the dispersion in tenure and age.

3.2.2 Validity of survey measures

Our survey measures have been tried-and-tested in earlier literature. Here, we report three internal validity statistics. First, we determine Cronbach's Alpha for all measures that consist of more than one question. Second, we report intra-class correlation coefficients, which provide an indication of the strength of within-store correlations in the survey measures. Third, we look at the consistency of survey measures over time, by looking at the correlation of measures between the pre-experiment and the post-experiment survey.

Column 1 of Table 5 gives the Cronbach's Alpha for all survey measures at the individual level, for both surveys separately. For social cohesion, the Cronbach's Alphas are comparable to those found in Carless and De Paola (2000) and indicate a sufficient degree of internal consistency in the scale. The same holds for the other measures, with the exception of directive leadership, which has a low Cronbach's Alpha. All Cronbach's Alpha's are slightly higher if we exclude the responses from the managers from calculating these measures.

Column 2 in Table 5 gives the intra-class correlation coefficients, a measure of the within-store correlation of survey outcomes across employees. The intra-class correlation of social cohesion is around 0.2. This is a moderate amount of within-store correlation.[21] Hence, within a team, employees' perception of the social cohesion in the team contains a common element, but also contains idiosyncratic elements. One possibility is that subsets of employees within stores do have good social relations with each-other, but that others are left out. For

---

[20]Obviously, the lack of correlation between social cohesion and sales growth does not rule out a positive correlation between social cohesion and the level of sales.

[21]In the pre-experiment (post-experiment) survey, store-fixed effects explain almost 40 (50) percent of variation across employees in social cohesion.

this reason, we include a robustness check in our analysis below. We create a measure of the dispersion of social cohesion across employees within a store, and analyze whether our results are affected if we control for this dispersion. For most other survey measures, the intra-class correlation is rather small as well. Hence, people within a store do vary in their perception of task cohesion, leadership, co-worker helping, and peer pressure.

Lastly, we assess the stability of the different survey measures over time by comparing the pre- and post-experiment survey outcomes. As the post-experiment measures in the treatment group may have been affected by the team incentive, we do so for stores in the control group only. Column 3 in Table 5 reports the correlations between the pre-experiment and the post-experiment survey measures at the individual employee level, while column 4 reports these correlations at the store level. As expected, the correlations at the individual level are generally higher than at the store level. For social cohesion, the correlation is 0.74 at the individual level and 0.61 at the store level, respectively. For the other survey measures, we also find high correlations, with the exception of directive leadership. This suggests that the survey measures do capture latent factors that are relatively stable over time.

Overall, our survey measures have reasonably good internal validity statistics, with the exception of directive leadership. None of our result are qualitatively affected if we drop directive leadership from our analysis.

3.2.3 Self-selection into the surveys

Given our design, employees' self-selection into the pre- and post-experiment surveys raises several possible concerns. If respondents are systematically different from non-respondents, our survey measures could be biased. This would be particularly problematic if self-selection into the surveys differs between employees from stores in the treatment group and employees from stores in the control group. Importantly, the decision to respond to the pre-experiment survey cannot be based on the assignment to treatment or control, as this survey was conducted before the stores' assignment. If assignment to treatment or control would affect employees' decision to respond to the post-experiment survey, our estimates of the effect of team incentives on the survey measures could be driven by differential self-selection of respondents rather than real changes within the stores. Fortunately, we find no indications of such differences in self-selection, as discussed below.

Column 1 of Table 6 shows that among managers, respondents are older, have longer tenure, and are more likely to be male than non-respondents.[22] This holds for both surveys. Among employees, there are no significant differences in personnel characteristics between

---

[22]Two stores did not have a store manager, implying that the total number of managers in our sample is 120.

Figure 5: Kernel densities of pre-exp. survey measures by response to post-exp. survey

Solid lines: Respondents did not respond to post-experiment survey.
Dashed lines: Respondents did respond to post-experiment survey.
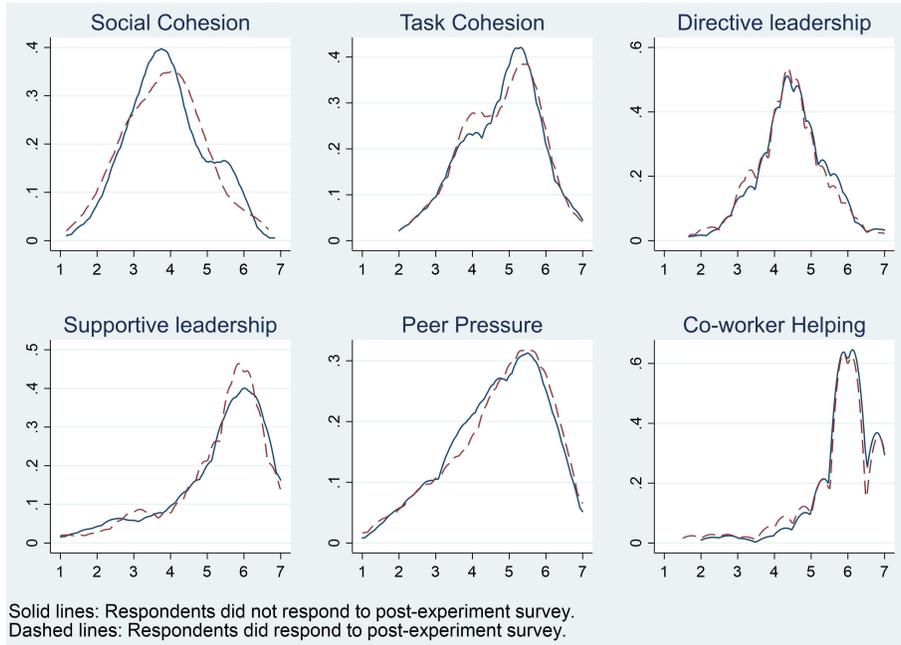
respondents and non-respondents in either survey. Columns 2 and 3 show that these findings also hold for the subsets of control stores and treatment stores. We find no significant differences between treatment and control in the differences in personnel characteristics between respondents and non-respondents. Response rates to both surveys are also similar for treatment and control.

Next, we compare the survey outcomes of the people that answered both surveys with the outcomes of the people that answered either only to the pre-experiment survey or only to the post-experiment survey.[23] Conditional on answering the pre-experiment survey, the probability of responding to the post-experiment survey is 62% in the treatment group and 57% in the control group. Similarly, conditional on responding to the post-experiment survey, the probability of having answered the pre-experiment survey is 87% in the treatment group and 88% in the control group. Neither of these differences between the treatment and the control group is statistically significant. Figure 5 shows the distributions of the survey measures in the pre-experiment survey, separately for people who did and did not answer the post-experiment survey. Similarly, Figure 6 gives the distributions of these survey measures in the post-experiment survey, for respondents who did and did not respond to the pre-experiment survey. Using Kolmogorov-Smirnov tests, we find no significant differences between the two groups for any measure in either survey.[24] Hence, self-selection into the

---

[23]In total, 285 employees completed both surveys, 189 completed the pre-experiment survey only, and 41 completed only the post-experiment survey.

[24]Performing these tests separately for respondents from stores in the treatment group and respondents

Figure 6: Kernel densities of post-exp. survey measures by response to pre-exp. survey

Solid lines: Respondents did not respond to pre-experiment survey.
Dashed lines: Respondents did respond to pre-experiment survey.
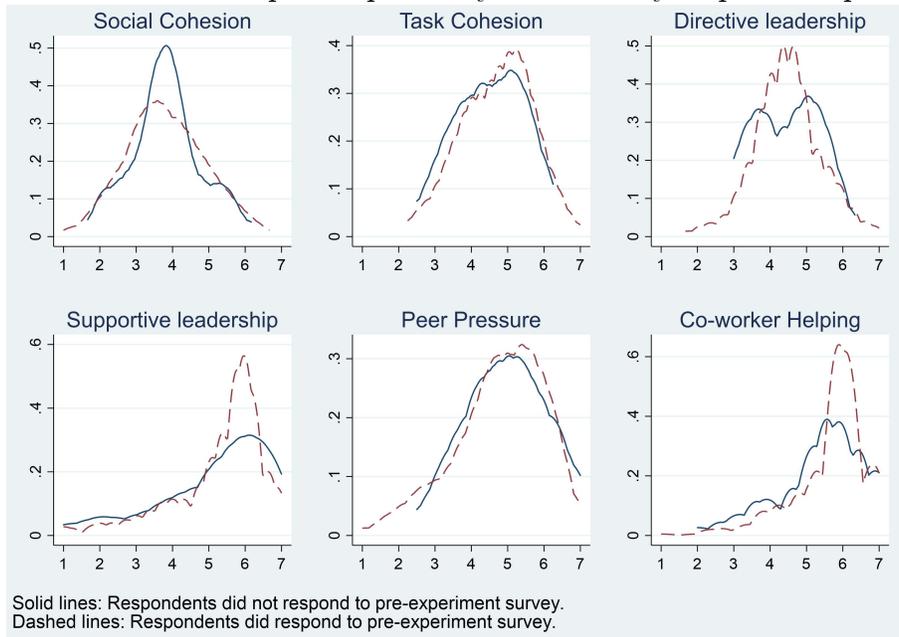
second survey appears unrelated to employees' responses to the first survey. Similarly, employees who only responded to the post-experiment survey do not provide different answers than employees who also answered the pre-experiment survey.

We cannot fully rule out a bias due to self-selection, as it may be driven by unobservables. For instance, if the decision to respond to the post-experiment survey depends on the change in a store's social cohesion during the experimental period, it is impossible to assess whether this differs between stores in the treatment group and stores in the control group. However, the findings discussed above show that on all observable dimensions, self-selection into the surveys appears unrelated to assignment to treatment or control.

4 Hypotheses and methodology

We use our experiment to test three hypotheses. First, we test whether the introduction of team incentives results in higher performance. This prediction follows from standard economic theory: the prospect of earning a bonus spurs additional effort from employees. We estimate the effect of the introduction of the team incentives on sales growth using panel OLS regressions with store- and week-fixed effects. Let $R_{s,w}$ denote the sales growth of store s in week w. The average treatment effect is estimated by:

$$R_{s,w} = \alpha_s + \gamma_w + \beta T_{s,w} + \varepsilon_{s,w} \tag{1}$$

from stores in the control group does not yield any statistically significant difference either.

where α_s and γ_w are store- and week-fixed effects, respectively. $T_{s,w}$ is a dummy that takes value 1 for stores in the treatment group during the experimental period, so that β captures the average treatment effect. $\varepsilon_{s,w}$ is the error term. In all estimations, we cluster standard errors at the store level, to adjust for (serial) correlation within stores (Bertrand et al. 2004).

Second, we test whether the response to team incentives is stronger in teams with more social cohesion. In theory, employees' response to team incentives is muted by the free-rider effect. The cost of effort is borne by the individual employee, but the benefits of effort are spread out across all team members. If employees are narrowly self-interested, their response to team incentives is sub-optimally low. However, if employees care about each other, the free-rider effect is mitigated as employees internalize (part of) the benefits of their effort that accrue to their team members (Rotemberg 1994; Dur and Sol 2010). Hence, we predict that social cohesion among employees in a store positively affects the store's response to team incentives.

Let C_s^q, with $q \in [1,2]$, denote the social cohesion of store s as measured in the pre-experiment questionnaire if $q = 1$ and in the post-experiment questionnaire if $q = 2$. We estimate the relation between store's social cohesion and the effect of team incentives on performance by interacting treatment dummy $T_{s,w}$ with stores' social cohesion as measured in the pre-experiment questionnaire C_s^1:

$$R_{s,w} = \alpha_s + \gamma_w + \beta T_{s,w} + \psi T_{s,w} C_s^1 + \phi E_{s,w} C_s^1 + \varepsilon_{s,w} \qquad (2)$$

Here, $E_{s,w}$ is a dummy variable that takes value 1 for all stores during the experimental period. Hence, through $E_{s,w} C_s^1$, we control for changes in the relation between social cohesion and sales growth that are unrelated to the experiment. Our main prediction is that $\psi > 0$: in stores with stronger social cohesion, performance increases more strongly in response to the introduction of team incentives as compared to stores where social cohesion is weaker. We will also perform a regression that allows for a nonlinear relation between social cohesion and the effect of team incentives on performance.

Stores' social cohesion C_s^1 is taken from the pre-experiment survey and, therefore, is not affected by our experimental introduction of team incentives. To deal with possible confounds, we control for the interactions between other store-level measures from the pre-experiment survey and the $T_{s,w}$ and $E_{s,w}$ dummies, to exclude that these drive the estimated relation between social cohesion and the response to treatment. Similarly, we control for interactions with store characteristics that may affect the response to team incentives, such as the number of employees in the team, the gender composition, and employees' tenure.[25]

---

[25] In Section 6, we perform a number of robustness checks. This includes a placebo treatment, where we perform the analysis as if the experiment took place in the six weeks prior to the actual experimental period.

Our third test concerns the effect of team incentives on social cohesion. Here, the existing literature provides competing hypotheses. Rotemberg (1994) and Dur and Sol (2010) argue that team incentives induce employees to invest in interpersonal relations. Better relations imply less free-riding. Hence, team incentives make it more rewarding to invest in co-worker relations, for instance by helping co-workers as in Fitzroy and Kraft (1986) and Drago and Turnbull (1988). Kandel and Lazear (1992) and Barron and Gjerde (1997), on the contrary, argue that stronger team incentives can lead to more peer pressure. As more is at stake, employees may coerce their colleagues to put in extra time and effort. This may lead to better performance, but also harm social relations among co-workers. Our design allows to determine which of these hypotheses best predicts the effect of team incentives on social cohesion in our context. To establish whether the mechanisms put forward in the literature drive the effect, we estimate the effects of team incentives on peer pressure, co-worker helping, and job satisfaction.

We choose to estimate the effects of team incentives on social cohesion, co-worker helping, peer pressure, and job satisfaction at the employee-level.[26] Let $C_{s,i}^q$ denote the social cohesion of employee i employed in store s as measured in questionnaire q. We estimate

$$C_{s,i}^q = \alpha_i + \gamma_q + \beta T_{s,q} + \varepsilon_{i,q} \tag{3}$$

where α_i and γ_q are individual and questionnaire-fixed effects, respectively. $T_{s,q}$ is the treatment dummy, which takes value 1 for individuals employed in the treatment stores for the post-experiment questionnaire ($q = 2$). Hence, β gives the average treatment effect of team incentives on social cohesion. $\varepsilon_{i,q}$ is an error term. Note that we again cluster errors at the store level. In the same way, we estimate the average treatment effect on the other survey measures.

5 Results

5.1 How does social cohesion affect the effect of team incentives on performance?

Column 1 in Table 7 presents the results of estimating the average treatment effect using regression equation (1). The average treatment effect of the team incentive is 1.5 percentage points higher sales growth. This effect is not statistically significant, but is well within the range of findings in comparable settings (see the Discussion section in Delfgaauw et al.

---

[26]As a robustness check, we also perform these estimations at the store level.

2019).[27]

In the absence of data on absolute sales revenue and margins, we cannot perform an exact cost-benefit analysis, but we can do an approximation. In the experiment, the employees of 18 stores won a bonus of, on average, 50 euro. As an average store has 10 employees, the total amount of bonus pay is about 9,000 euro. Furthermore, the company spend about 700 euro on communication (sending letters and posters), implying a total cost of approximately 10,000 euro. Using figures from Statistics Netherlands, the average shoe shop in The Netherlands earns about 600,000 euro per year, with a gross margin of about 45 percent. As the retail chain where we conducted our study targets budget-minded consumers, we use conservative estimates to approximate the effect of our intervention on revenue. Our experiment ran for 6 weeks in 72 stores. Using yearly revenue of 500,00 euro per store and a gross margin of 30 percent, this implies that an increase in sales during the experiment of about 0.8 percentage point would suffice for breaking even. The estimates in column 1 of Table 7 imply that the point estimate of the effect on profits is about 8,700 euro, and the probability that the experiment increased profit equals 72 percent.

Column 2 in Table 7 gives the results of estimating (2). In line with the theoretical prediction, we find that the effect of the team incentive on sales growth increases in stores' social cohesion. Our estimation of ψ in equation (2) equals 2.891. This implies a sizable relation between social cohesion and the response to team incentives: An increase in social cohesion by one standard deviation (0.79 on a 7-point scale) corresponds to an increase in the predicted effect of team incentives on sales by 2.3 percentage points.

In Figure 7, we plot the estimated treatment effect as a function of social cohesion, assuming a linear-quadratic relation between social cohesion and the response to team incentives. Here, we find that the team incentive hardly affects sales in stores with low to moderate social cohesion. If social cohesion is sufficiently high, the team incentive positively affects sales growth. The estimated effect of team incentives on sales growth is statistically significant for about 25% of the stores.

Column 3 of Table 7 adds interaction effects for task cohesion and leadership. The estimated interaction effect of team incentives and social cohesion is hardly affected. Furthermore, task cohesion and employee-manager relations do not interact with the response to team incentives. Hence, even though social cohesion, task cohesion, and leadership styles are correlated, only social cohesion is related to stores' response to team incentives.

In column 4, we also control for other store characteristics that may affect the response

---

[27]The absence of a statistically significant effect is not due to a lack of power: We can detect an effect size of 3 percentage points with 80% probability. This is derived as follows. We perform a regression with store- and week-fixed effects. In the six-week period of the experiment, the residuals from this regression have a standard deviation of 11.56 and an intra-cluster correlation coefficient of 0.095. Using standard power calculations, this yields a minimal detectable effect size of 2.94 percentage points.

Figure 7: Estimated treatment effect as a function of social cohesion

The gray area depicts the density of social cohesion across stores.
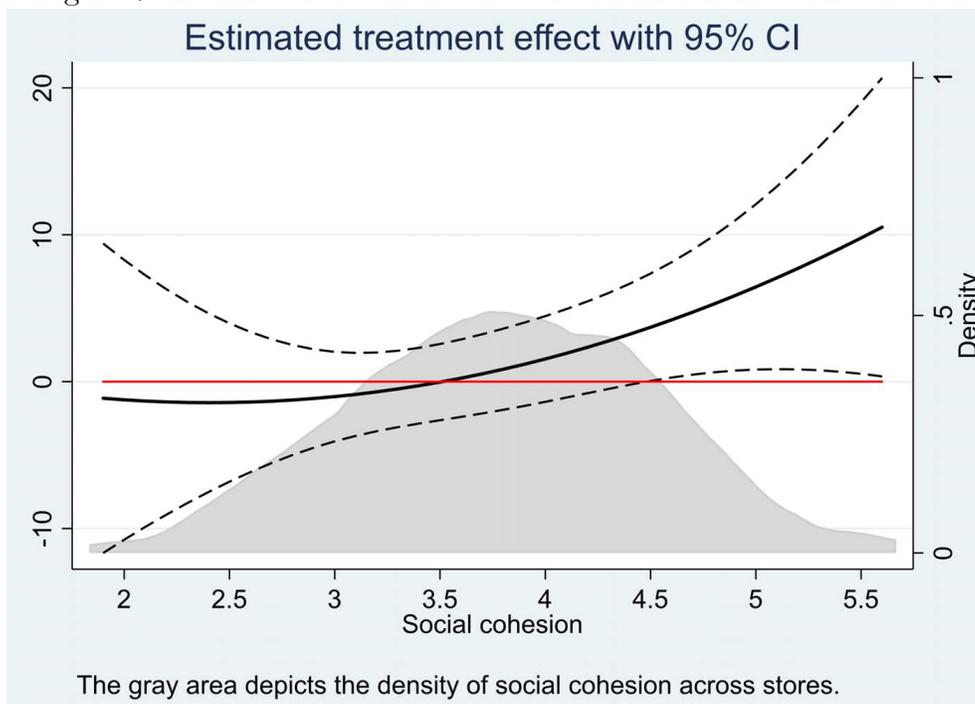
to team incentives: The number of employees and the dispersion in age, tenure, and gender among employees, where dispersion is measured by the standard deviation within the store. Again, we find that the estimated interaction effect between social cohesion and team incentives is not affected, although the standard errors are slightly higher. None of the store characteristics relates significantly to the response to team incentives.

As a final robustness check, we regressed social cohesion on task cohesion, leadership styles, and all store characteristics in our data, see Table A1 in the Appendix. Next, we interacted the residuals of this regression with the treatment dummy. The store-level variables explain about 29 percent of the variation in social cohesion across stores. We find that the interaction effect of the residuals of the social cohesion regression and the treatment has the same magnitude as the estimations reported in Table 7: a coefficient of 3.17 with a standard error of 1.91. Hence, we conclude that the positive relation between social cohesion and the effect of team incentives on performance as reported in column 2 of Table 5 is not driven by observable factors.

5.2 How do team incentives affect social cohesion?

Panel A in Table 8 gives the results of estimating (3). We find a small and statistically insignificant effect of team incentives on social cohesion. The point estimate of 0.04 on a 7-point scale corresponds to an increase of 3.7 percent of a standard deviation in social

cohesion. Given the small standard error of the estimate, we can rule out that the team incentive had a sizable effect on social cohesion. Hence, the introduction of team incentives neither hurt nor improved social relations among employees.

We estimate a negative and statistically insignificant effect of team incentives on peer pressure. We do find a statistically significant positive effect on co-worker helping, but the estimated effect size is still quite small: 0.23 on a 7 point scale, corresponding to 21 percent of a standard deviation in co-worker helping. The estimated effect on job satisfaction is essentially zero.[28]

We find similar results if we exclude the store managers from these estimations. Furthermore, if we perform the estimations at the store level, the effect on co-worker helping is smaller and statistically insignificant. All other findings are qualitatively similar.[29]

Panel B in Table 8 shows that there is no statistically significant treatment heterogeneity with respect to pre-existing social cohesion. Hence, the small average treatment effects reported in Panel A do not mask differences in response to team incentives between teams that are more and less socially cohesive.[30] Again, using store-level measures rather than individual-level measures does not affect these findings qualitatively.

Overall, the effects of the team incentives on social interactions and on job satisfaction within the stores are limited, if any. Hence, the team incentives neither improved nor harmed the atmosphere within stores. This may comfort managers who consider introducing team incentives but worry about the effects on social interactions within the team. At the same time, our results suggest that team incentives are no panacea for bad co-worker relations.

6 Robustness checks

In this section, we discuss a number of robustness checks. First, we perform a placebo treatment, where we pretend that the treatment took place in the six weeks before we implemented the actual experiment (weeks 36 - 41 instead of 42 - 47). Columns 1 and 2 in Table 9 report the estimated effects of the placebo treatment. Both the average placebo treatment effect and the interaction between the placebo treatment and social cohesion are considerably smaller than the corresponding estimates of the actual treatment (Table 7, columns 1 and 2) and not statistically significant. Note that the coefficient on "Exp. period

---

[28]This also holds for the effects on job search and intention to quit. For exploratory reasons, we also estimated the treatment effect on task cohesion and leadership styles. Again, the estimated effects are small and not statistically significant.

[29]Non-reported estimations show that these average treatment effects do not hide substantial differences between stores that did and did not win their competition.

[30]Mean-reversion is the most likely explanation for the negative relation between measured social cohesion in the pre- and the post-experiment questionnaires.

X Social Cohesion" has the same magnitude as in the main estimations. Hence, the negative relation between social cohesion and sales growth during the experimental period is part of a pre-existing trend, ruling out that contamination drives this relation.

Second, as an alternative for clustering standard errors at the store level to correct for serial correlation, we perform a before - after difference-in-differences estimation (Bertrand et al. 2004). Hence, we collapse all observations per store into two observations: average sales growth before the experimental period and average sales growth during the experimental period. Columns 3 and 4 of Table 9 show that the results of this difference-in-differences estimation are very close to the results of our main specification as reported in columns 1 and 2 of Table 7. The standard error on the interaction effect between the treatment dummy and store's pre-experiment social cohesion increases just enough to make the effect statistically insignificant at the 10% level.

Third, it is possible that the intermediate standings during the competition affected stores' response. We do not attempt to estimate these dynamic effects.[31] If stores respond to the rankings provided during the experiment, the error terms of stores within a competition group may be correlated. To correct for this, we have repeated our estimations of (1) and (2) while clustering standard errors at the competition group level for treatment stores. For control stores, clustering remains at the store level, implying a total of 68 clusters. This does not meaningfully affect the standard errors in our estimations, as reported in columns 5 and 6 in Table 9.

Fourth, we create a measure of stores' social cohesion that exclude the responses of the store managers. Arguably, their management role could give them a distinct position within the team of employees or may give them a different view on the team's social interactions. However, we find little evidence that supports this. The correlation between the average level of social cohesion as reported by a store's employees (excluding the manager) and the level of social cohesion as reported by the manager is 0.33. Replacing our measure of social cohesion with the measure that excludes the responses by managers hardly affects the estimated interaction effect between the introduction of team incentives and stores' social cohesion, as shown in column 7 of Table 9.

Lastly, as discussed in Section 3.2, there is variation in reported social cohesion within stores. This may imply that people within a store also react differentially to team incentives. Lacking individual measures of performance, we cannot directly assess this. Instead, we analyze whether controlling for the interaction with the within-store dispersion in reported social cohesion affects our findings. Thereto, we use the within-store standard deviation

---

[31]As intermediate standings are not exogenous, estimating their effects is complicated in the presence of serial correlation. Delfgaauw et al. (2014) implements a design that allows for estimating these dynamic effects and finds modest dynamic effects.

of social cohesion at the individual level as the measure for dispersion.[32] Column 8 of Table 9 shows that controlling for dispersion hardly affects the estimated relation between social cohesion and the response to team incentives, but does increase its standard error. The within-store dispersion in social cohesion itself does not have a statistically significant relation with the response to team incentive.[33]

7 Concluding remarks

By combining rich questionnaire data and a large-scale experiment on incentive pay in an actual firm, we have studied the interplay between team incentives, social cohesion, and team performance. We find that team performance increases more after the introduction of team incentives if team members have better social relations. This result is in line with the theoretical prediction that in teams with good social relations, individuals incorporate the external effects of their efforts on others and, hence, engage in less free-riding. This result also exemplifies the importance of complementarities in organizations. Firms that introduce or reinforce team incentives would benefit from enhancing social interactions among employees.

The effect of team incentives on social cohesion within teams turns out to be small. Hence, team incentives do not instantly affect social interactions between colleagues. One caveat here is the relatively short incentive period. Perhaps six weeks is too short to create lasting changes in social relations among employees in real-world workplaces. Whether and how long-lasting changes in incentive pay affect social interactions remains an open question. More generally, given that both firms and employees value good social relations highly, establishing how management practices affect social cohesion at work is an important area for future research.

---

[32] For 6 stores, we have only one respondent in the pre-experiment questionnaire, implying that the standard deviation cannot be computed. We drop these stores in the reported estimation. Our findings are qualitatively similar if, instead, we create a dummy variable for stores with only one respondent.

[33] For all robustness checks except the placebo treatment, we find that a quadratic specification of the interaction between social cohesion and the introduction of the team incentives yields estimated effects close to those depicted in Figure 7.

References

[1] Akerlof, George A., and Rachel E. Kranton (2000), Economics and Identity, *Quarterly Journal of Economics* 115(3): 715 – 753.

[2] Ashraf, Nava, and Oriana Bandiera (2018), Social Incentives in Organizations, *Annual Review of Economics* 10: 439 – 463.

[3] Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman, and Heather Royer (2015), Letting Down the Team? Social Effects of Team Incentives, *Journal of the European Economic Association* 13(5): 841 – 870.

[4] Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2005), Social Preferences and the Response to Incentives: Evidence from Personnel Data, *Quarterly Journal of Economics* 120(3): 917 – 962.

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

[6] 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 – 1114.

[7] Barron, John M., and Kathy Paulson Gjerde (1997), Peer Pressure in an Agency Relationship, *Journal of Labor Economics* 15(2): 234 – 254.

[8] Beal, Daniel J., Robin R. Cohen, Michael J. Burke, and Christy L. McLendon (2003), Cohesion and Performance in Groups: A Meta-Analytic Clarification of Construct Relations, *Journal of Applied Psychology* 88(6): 989 – 1004.

[9] Bell, Suzanne T. (2007), Deep-level Composition Variables as Predictors of Team Performance: A Meta-analysis, *Journal of Applied Psychology* 92(3): 595 – 615.

[10] Berger, Johannes, Claus Herbertz, and Dirk Sliwka (2011), Incentives and Cooperation in Firms: Field Evidence, IZA DP No. 5618.

[11] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), How Much Should We Trust Differences-in-Differences Estimates?, *Quarterly Journal of Economics* 119(1): 249 – 275.

[12] Bloom, Nicholas, and John Van Reenen (2007), Measuring and Explaining Management Practices Across Firms and Countries, *Quarterly Journal of Economics* 122(4): 1351 – 1408.

[13] 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.

[14] Böckerman, Petri, and Pekka Ilmakunnas (2009), Job Disamenities, Job Satisfaction, Quit Intentions, and Actual Separations: Putting the Pieces Together, *Industrial Relations* 48(1): 73 – 96.

[15] 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 – 989.

[16] Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez (2012), Inequality at Work: The Effect of Peer Salaries on Job Satisfaction, *American Economic Review* 102(6): 2981 – 3003.

[17] Carless, Sally A., and Caroline De Paola (2000), The Measurement of Cohesion in Work Teams, *Small Group Research* 31(1): 71 – 88.

[18] Carpenter, Jeffrey, and Erika Seki (2011), Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyoma Bay, *Economic Inquiry* 49: 612 – 630.

[19] Carron, Albert V., and Kevin S. Spink (1995), The Group Size-Cohesion Relationship in Minimal Groups, *Small Group Research* 26(1): 86 – 105.

[20] Casey-Campbell, Milly, and Martin L. Martens (2009), Sticking It All Together: A Critical Assessment of the Group Cohesion – Performance Literature, *International Journal of Management Reviews* 11(2): 223 – 246.

[21] Charness, Gary, Luca Rigotti, and Aldo Rustichini (2007), Individual Behavior and Group Membership, *American Economic Review* 97(4): 1340 – 1352.

[22] Chen, Yan, and Sherry Xin Li (2009), Group Identity and Social Preferences, *American Economic Review* 99(1): 431 – 457.

[23] Chen, Hua, and Noah Lim (2013), Should Managers Use Team-Based Contests?, *Management Science* 59(12): 2823 – 2836.

[24] Chiocchio, François, and Hélène Essiembre (2009), Cohesion and Performance: A Meta-Analytic Review of Disparities Between Project Teams, Production Teams, and Service Teams, *Small Group Research* 40: 382 – 420.

[25] Clark, Andrew E. (2001), What Really Matters in a Job? Hedonic Measurement Using Quit Data, *Labour Economics* 8: 223 – 242.

[26] Corgnet, Brice, Roberto Hernán-Gonzalez, and Ricardo Mateo (2019), Rac(g)e Against the Machine? Social Incentives When Humans Meet Robots, mimeo, Chapman University.

[27] Cornelissen, Thomas (2009), The Interaction of Job Satisfaction, Job Search, and Job Changes. An Empirical Investigation with German Panel Data, *Journal of Happiness Studies* 10(3): 367 – 384.

[28] Delfgaauw, Josse, Robert Dur, Okemena A. Onemu, Joeri Sol, and Willem Verbeke (2011), A Field Experiment on Team Incentives and Peer Pressure, In: Joeri Sol, Incentives and Social Relations in the Workplace, 61 – 85, Tinbergen Institute PhD Thesis.

[29] 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 – 326.

[30] 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.

[31] Delfgaauw, Josse, Robert Dur, Arjan Non, and Willem Verbeke (2015), The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment, *Journal of Labor Economics* 33(3): 521 – 569.

[32] Delfgaauw, Josse, Robert Dur, and Michiel Souverijn (2019), Team Incentives, Task Assignment, and Performance: A Field Experiment, *The Leadership Quarterly* forthcoming.

[33] Deloitte (2016), Global Human Capital Trends: The New Organizational: Different by Design. London: Deloitte University Press.

[34] De Paola, Maria, Francesca Gioia, and Vincenzo Scoppa (2019), Free-riding and Knowledge Spillovers in Teams: The Role of Social Ties, *European Economic Review* 112: 74 – 90.

[35] Dion, Kenneth L. (2000), Group Cohesion: From "Field of Forces" to Multidimensional Construct, *Group Dynamics: Theory, Research, and Practice* 4(1): 7 – 26.

[36] Drago, Robert, and Geoffrey K. Turnbull (1988), Individual versus Group Piece Rates under Team Technologies, *Journal of the Japanese and International Economies* 2(1): 1 – 10.

[37] Dur, Robert, and Joeri Sol (2010), Social Interaction, Co-Worker Altruism, and Incentives, *Games and Economic Behavior* 69(2): 293 – 301.

[38] Englmaier, Florian, Andreas Roider, and Uwe Sunde (2016), The Role of Communication of Performance Schemes: Evidence from a Field Experiment, *Management Science* 63(12): 4061 – 4080.

[39] Englmaier, Florian, Stefan Grimm, David Schindler, and Simeon Schudy (2018), The Effect of Incentives in Non-Routine Analytical Teams Tasks: Evidence from a Field Experiment, CESifo working paper no. 6903.

[40] Erev, Ido, Gary Bornstein, and Rachely Galili (1993), Constructive Intergroup Competition as a Solution to the Free Rider Problem: A Field Experiment, *Journal of Experimental Social Psychology* 29(6): 463 – 478.

[41] Festinger, Leon (1950), Informal Social Communication, *Psychological Review* 57: 271 – 282.

[42] FitzRoy, Felix R., and Kornelius Kraft (1986), Profitability and Profit-Sharing, *Journal of Industrial Economics* 35(2): 113 – 130.

[43] Freeman, Richard B., Douglas L. Kruse, and Joseph R. Blasi (2008), The Same Yet Different: Worker Reports on Labor Practices in a Single Firm Across Countries, *Labour Economics* 15(4): 749 – 770.

[44] Freeman, Richard B., Douglas L. Kruse, and Joseph R. Blasi (2010), Worker Responses to Shirking under Shared Capitalism, In: Douglas L. Kruse, Richard B. Freeman and Joseph R. Blasi (Editors), *Shared Capitalism at Work: Employee Ownership, Profit and Gain Sharing, and Broad-based Stock Options* 77 – 103, University of Chicago Press.

[45] Friebel, Guido, Matthias Heinz, Miriam Krueger, and Nikolay Zubanov (2017), Team Incentives and Performance: Evidence from a Retail Chain, *American Economic Review* 107(8): 2168 – 2203.

[46] Friedkin, Noah E. (2004), Social Cohesion, *Annual Review of Sociology* 30: 409 – 425.

[47] Fryer, Roland (2013), Teacher Incentives and Student Achievement: Evidence from New York City Public Schools, *Journal of Labor Economics* 31(2): 373 – 407.

[48] Gallup (2017), State of the American Workplace, Washington: Gallup inc.

[49] Glewwe, Paul, Nauman Ilias, and Michael Kremer (2010), Teacher Incentives, *American Economic Journal: Applied Economics* 2(3): 205 – 227.

[50] Goette, Lorenz, David Huffman, Stephan Meier, and Matthias Sutter (2012), Competition Between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations, *Management Science* 58(5): 948 – 960.

[51] Goette, Lorenz, David Huffman, and Stephan Meier (2006), The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups, *American Economic Review* 96(2): 212 – 216.

[52] Green, Jerry R., and Nancy L. Stokey (1983), A Comparison of Tournaments and Contracts, *Journal of Political Economy*91(3): 349 – 364.

[53] Griffeth, Rodger W., Peter W. Hom, and Stefan Gaertner (2000), A Meta-analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the Next Millennium, *Journal of Management* 26(3): 463 – 488.

[54] Gully, Stanley M., Dennis J. Devine, David J. Whitney (1995), A Meta-analysis of Cohesion and Performance: Effects of Levels of Analysis and Task Interdependence, *Small Group Research* 20: 70 – 86.

[55] 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 – 497.

[56] Harrison, David A., Kenneth H. Price, Joanne H. Gavin, and Anna T. Florey (2002), Time, Teams, and Task Performance: Changing Effects of Surface- and Deep-Level Diversity on Group Functioning, *Academy of Management Journal* 45(5): 1029 – 1045.

[57] Harrison, Glenn, and John A. List (2004), Field Experiments, *Journal of Economic Literature* 42(4): 1009 – 1055.

[58] Holmström, Bengt (1982), Moral Hazard in Teams, *Bell Journal of Economics* 13(2): 324 – 340.

[59] Kandel, Eugene, and Edward P. Lazear (1992), Peer Pressure and Partnerships, *Journal of Political Economy* 100(4): 801 – 817.

[60] Keller, Robert T. (1983), Predicting Absenteeism From Prior Absenteeism, Attitudinal Factors, and Nonattitudinal Factors, *Journal of Applied Psychology* 68(3): 536 – 540.

[61] Lavy, Victor (2002), Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement, *Journal of Political Economy* 110(6): 1286 – 1317.

[62] Lazear, Edward P. (2018), Compensation and Incentives in the Workplace, *Journal of Economic Perspectives* 32(3): 195 – 214.

[63] Lazear, Edward P., and Sherwin Rosen (1981), Rank-Order Tournaments as Optimum Labor Contracts, *Journal of Political Economy* 89(5): 841 – 864.

[64] Levitt, Steven D., and Susanne Neckermann (2014), What Field Experiments Have and Have Not Taught Us About Managing Workers, *Oxford Review of Economic Policy* 30(4): 639– 657.

[65] Lim, Noah, and Hua Chen (2014), When Do Group Incentives for Salespeople Work? *Journal of Marketing Research* 51(3): 320 – 334.

[66] List, John A., Sally Sadoff, and Mathis Wagner (2011), So You Want To Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design, *Experimental Economics* 14, 439 – 457.

[67] McKinsey (2016), How Social Tools Can Reshape the Organization, McKinsey & Company.

[68] McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001), Birds of a Feather: Homophily in Social Networks, *Annual Review of Sociology* 27(1): 415 – 444.

[69] Moorman, Robert H., and Gerald L. Blakely (1995), Individualism-Collectivism as an Individual Difference Predictor of Organizational Citizenship Behavior, *Journal of Organizational Behavior* 16: 127 – 142.

[70] Mullen, Brian, and Carolyn Copper (1994), The Relation Between Group Cohesiveness and Performance: An Integration, *Psychological Bulletin* 155(2): 210 – 227.

[71] Mueller, Charles W., and James L. Price (1990), Economic, Psychological, and Sociological Determinants of Voluntary Turnover, *Journal of Behavioral Economics* 19(3): 321– 335.

[72] Muralidharan, Karthik, and Venkatesh Sundararaman (2011), Teacher Performance Pay: Experimental Evidence from India, *Journal of Political Economy* 119(1): 39–77.

[73] Nalebuff, Barry J., and Joseph E. Stiglitz (1983), Prizes and Incentives: Towards a General Theory of Compensation and Competition, *Bell Journal of Economics* 14(1): 21 – 43.

[74] Nishii, Lisa H., and David M. Mayer (2009), Do Inclusive Leaders Help to Reduce Turnover in Diverse Groups? The Moderating Role of Leader–Member Exchange in the Diversity to Turnover Relationship, *Journal of Applied Psychology* 94(6): 1412 – 1426.

[75] Price, James L., and Charles W. Mueller (1986), *Absenteeism and Turnover of Hospital Employees*, JAI Press.

[76] Rotemberg, Julio J. (1994), Human Relations in the Workplace, *Journal of Political Economy* 102(4): 684 – 717.

[77] Sheremeta, Roman M. (2018), Behavior in Group Contests: A Review of Experimental Research, *Journal of Economic Surveys* 32(3): 683 – 704.

[78] Tajfel, Henri, and John C. Turner (1979), An Integrative Theory of Intergroup Conflict," In William G. Austin and Stephen Worchel, eds., *The Psychology of Intergroup Relations.* Monteray: Brooks/Cole Publishing Co., 33 – 47.

[79] Tett, Robert P., and John P. Meyer (1993), Job Satisfaction, Organizational Commitment, Turnover Intention, and Turnover: Path Analyses based on Meta-analytic Findings, *Personnel Psychology* 46: 259 – 293.

[80] Waber, Ben, Jennifer Magnolfi, and Greg Lindsay (2014), Workspaces That Move People, *Harvard Business Review* October.

[81] Wendt, Hein, Martin C. Euwema, and Hetty van Emmerik (2009), Leadership and Team Cohesiveness across Cultures, *The Leadership Quarterly* 20: 358 – 370.

[82] Widmeyer, W. Neil, Lawrence R. Brawley, and Albert V. Carron, (1985). The Measurement of Cohesion in Sports Teams: The Group Environment Questionnaire. London, Ontario: Sports Dynamics.

[83] Williams, Katherine Y., and Charles A. O'Reilly III (1998), Demography and Diversity in Organizations: A Review of 40 Years of Research, *Research in Organizational Behavior* 20: 77 – 140.

[84] Zaccaro, Stephen J., and Charless A. Lowe (1988), Cohesiveness and Performance on an Additive Task: Evidence for Multidimensionality, *Journal of Social Psychology* 128(4): 547 – 558.

[85] Zaccaro, Stephen J., and M. Catherine McCoy (1988), The Effects of Task and Interpersonal Cohesiveness on Performance of a Disjunctive Group Task, *Journal of Applied Social Psychology* 18(10): 837 – 851.

Table 1 is displayed in the main text on page 9.

Table 2: Descriptive statistics

	All stores		Control group		Treatment group	
	Mean	Std	Mean	Std	Mean	Std
Performance (store averages)						
Avg weekly sales growth before experimental period (week 1 - week 41 in 2013)	0.26	9.01	-0.88	8.73	1.05	9.18
Avg weekly sales growth during experimental period (week 42 - week 47 in 2013)	-6.38	11.34	-8.41	10.71	-4.96 *	11.62
Within-store standard deviation of sales growth before experimental period	30.34	7.42	30.66	8.14	30.12	6.93
Store characteristics						
Number of employees	9.87	2.71	9.78	2.45	9.93	2.88
Fraction male employees	0.15	0.13	0.15	0.14	0.14	0.12
Avg tenure of employees (years)	5.99	2.60	6.29	2.77	5.79	2.49
Avg age of employees	27.22	4.42	27.21	4.42	27.24	4.46
Fraction of on-call employees	0.41	0.14	0.42	0.13	0.40	0.14
Gender store manager (male = 1, female = 0)	0.58	0.49	0.56	0.49	0.59	0.49
Tenure store manager (years)	13.98	8.71	14.11	8.01	13.89	9.23
Age store manager	39.98	9.28	40.77	9.51	39.43	9.15
Number of stores	122		50		72	

The personnel variables are extracted from the company's database as of September 2013.
***, **, * denote statistically significant differences at the 1%, 5%, and 10% level, respectively, between control and treatment stores (t-test).

Table 3: Survey data descriptives

	All stores		Control group		Treatment group	
	Mean	Std	Mean	Std	Mean	Std
Panel A: Pre-experiment survey						
Response rate employees	0.35	0.18	0.34	0.17	0.36	0.18
Response manager	0.79	0.40	0.80	0.40	0.79	0.41
Store-level measures						
Social cohesion	3.80	0.71	3.79	0.77	3.81	0.68
Task cohesion	4.79	0.68	4.70	0.62	4.86	0.71
Directive leadership	4.48	0.57	4.43	0.52	4.51	0.60
Supportive leadership	5.31	0.87	5.28	0.86	5.32	0.88
Helping	5.86	0.60	5.84	0.55	5.87	0.64
Peer pressure	4.82	0.73	4.78	0.69	4.84	0.77
Job satisfaction	5.71	0.68	5.65	0.69	5.76	0.67
Number of stores with responses	120		50		70	
Panel B: Post-experiment survey						
Response rate employees	0.23	0.16	0.20	0.15	0.24	0.17
Response manager	0.68	0.47	0.70	0.46	0.67	0.47
Store-level measures						
Social cohesion	3.72	0.79	3.58	0.84	3.83	0.75
Task cohesion	4.68	0.71	4.65	0.80	4.70	0.65
Directive leadership	4.46	0.70	4.40	0.84	4.51	0.57
Supportive leadership	5.27	1.02	5.23	1.13	5.30	0.95
Helping	5.67	0.71	5.61	0.69	5.72	0.72
Peer pressure	4.79	1.00	4.83	1.06	4.76	0.96
Job satisfaction	5.63	0.96	5.54	1.06	5.70	0.89
Number of stores with responses	114		48		66	

***, **, * denote statistically significant differences at the 1%, 5%, and 10% level, respectively, between control and treatment stores (t-test).
All survey measures are on a 7-point Likert scale.

Table 4: Correlations

	Avg sales growth before exp.	Socal cohesion	Task cohesion	Directive leadership	Supportive leadership	Helping	Peer pressure	Number of employees	Average tenure employees	Average age employees	% male employees	% on-call employees	std tenure employees
Avg sales growth before experiment	1.00												
Socal cohesion	-0.04	1.00											
Task cohesion	-0.04	0.31	1.00										
Directive leadership	-0.11	-0.14	-0.07	1.00									
Supportive leadership	0.14	0.34	0.30	-0.18	1.00								
Helping	-0.02	0.39	0.64	-0.15	0.47	1.00							
Peer pressure	-0.05	-0.29	-0.15	0.04	0.00	-0.04	1.00						
Number of employees	0.05	0.12	-0.03	-0.05	-0.12	-0.06	-0.06	1.00					
Average tenure employees	-0.11	0.06	-0.17	0.06	0.17	-0.20	-0.09	0.01	1.00				
Average age employees	-0.04	-0.04	-0.11	0.01	0.11	-0.19	-0.04	-0.06	0.74	1.00			
% male employees	-0.07	0.00	0.07	-0.06	-0.01	0.05	-0.09	0.11	-0.14	-0.04	1.00		
% on-call employees	0.01	-0.02	0.08	-0.02	-0.14	0.06	-0.10	0.51	-0.36	-0.44	0.14	1.00	
std tenure employees	-0.17	-0.03	-0.20	0.14	0.00	-0.22	-0.02	0.05	0.72	0.47	-0.03	-0.20	1.00
std age employees	0.04	-0.18	-0.16	0.00	0.03	-0.14	-0.11	-0.01	0.49	0.64	-0.14	-0.07	0.53

The personnel variables are extracted from the company's database as of September 2013. The survey measures are derived from the pre-experiment survey.

All correlations are based on 120 stores, except for peer pressure, which is based on 117 stores.

Table 5: Validity of survey measures

	(1) Cronbach's Alpha	(2) Intra-class correlation coefficient	(3) Correlation with post- experiment survey, individuals	(4) Correlation with post- experiment survey, stores
Panel A: Pre-experiment survey				
Social cohesion	0.72	0.18	0.74	0.61
Task cohesion	0.58	0.10	0.51	0.36
Directive leadership	0.41	0.06	0.33	0.11
Supportive leadership	0.88	0.18	0.67	0.45
Helping	0.81	0.02	0.57	0.51
Peer pressure	0.66	0.00	0.62	0.46
Job satisfaction		0.04	0.78	0.55
N	430 - 474	430 - 474	96 - 107	47 - 48
Panel B: Post-experiment survey				
Social cohesion	0.75	0.22		
Task cohesion	0.49	0.12		
Directive leadership	0.26	0.11		
Supportive leadership	0.88	0.18		
Helping	0.77	0.00		
Peer pressure	0.66	0.07		
Job satisfaction		0.12		
N	302 - 326	302 - 326		

Table 6: Respondent vs non-respondents

	All stores				Control group				Treatment group			
	Respondents		Non-respondents		Respondents		Non-respondents		Respondents		Non-respondents	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
First Questionnaire												
Managers												
Number	94		26		39		11		55		15	
Age	40.66 **	9.26	36.50	9.04	41.49	9.68	36.55	8.29	40.07	8.99	36.47	9.85
Tenure	14.77 **	8.76	10.28	7.97	14.49	7.95	11.12	8.64	14.96 **	9.36	9.66	7.69
Male	0.62 *	0.49	0.42	0.50	0.62	0.49	0.36	0.50	0.62	0.49	0.47	0.52
Employees												
Number	380		704		149		290		231		414	
Age	26.89	11.04	27.32	12.14	26.96	11.45	27.20	11.94	26.84	10.79	27.41	12.29
Tenure	5.74	7.03	6.15	7.36	6.23	7.51	6.32	7.80	5.42	6.71	6.03	7.12
Male	0.14	0.35	0.16	0.36	0.17	0.37	0.17	0.38	0.12	0.33	0.15	0.36
On-call	0.40	0.49	0.44	0.50	0.42	0.50	0.44	0.50	0.38	0.49	0.44	0.50
Second Questionnaire												
Managers												
Number	80		40		34		16		46		24	
Age	41.29 **	9.65	36.70	7.94	42.38	10.17	36.19	6.47	40.48	9.27	37.04	8.90
Tenure	15.14 **	8.97	11.12	7.76	15.36	8.12	10.33	7.28	14.97	9.63	11.64	8.17
Male	0.63	0.49	0.48	0.51	0.62	0.49	0.44	0.51	0.63	0.49	0.50	0.51
Employees												
Number	246		838		88		351		158		487	
Age	26.80	10.97	27.28	11.99	27.48	11.93	27.03	11.73	26.43	10.42	27.46	12.18
Tenure	5.89	7.24	6.04	7.25	6.60	7.78	6.21	7.60	5.50	6.92	5.91	7.00
Male	0.12 *	0.32	0.16	0.37	0.15	0.36	0.17	0.38	0.10	0.30	0.15	0.36
On-call	0.42	0.49	0.43	0.50	0.43	0.50	0.44	0.50	0.41	0.49	0.43	0.49

The personnel variables are extracted from the company's database as of October 2013.

***, **, * denote statistically significant differences at the 1%, 5%, and 10% level, respectively, between respondents and non-respondents within
a questionnaire (t-test, non-parametric (Wilcoxon rank-sum) test yields same results). None of the differences between respondents and non-respondents
are significantly different between the control group and the treatment group at the 10% level.

Table 7: Treatment effects on sales growth

	Dependent variable: sales growth (percentage points)			
	(1)	(2)	(3)	(4)
Treatment	1.495	1.628	1.410	1.356
	(1.221)	(1.210)	(1.202)	(1.162)
Treatment X Social Cohesion		2.891*	3.239*	3.571*
		(1.686)	(1.844)	(1.964)
Exp. period X Social Cohesion		-2.365*	-2.755**	-2.900**
		(1.279)	(1.324)	(1.436)
Treatment X Task Cohesion			-0.926	-1.253
			(1.930)	(1.966)
Exp. period X Task Cohesion			1.746	1.756
			(1.423)	(1.425)
Treatment X Directive leadership			-0.880	-0.415
			(2.265)	(2.263)
Exp. period X Directive leadership			0.357	0.398
			(1.819)	(1.815)
Treatment X Supportive leadership			-1.180	-1.024
			(1.552)	(1.535)
Exp. period X Supportive leadership			0.536	0.555
			(1.161)	(1.222)
Treatment X Number of employees				-0.293
				(0.495)
Exp. period X Number of employees				0.547
				(0.441)
Treatment X std tenure				-0.845
				(0.560)
Exp. period X std tenure				0.143
				(0.402)
Treatment X std age				0.422
				(0.461)
Exp. period X std age				0.090
				(0.314)
Treatment X std males				-0.069
				(8.851)
Exp. period X std males				-1.209
				(7.502)
Store-fixed effects	yes	yes	yes	yes
Week-fixed effects	yes	yes	yes	yes
Store-week observations	5722	5628	5628	5628
Stores	122	120	120	120
R^2	0.638	0.638	0.638	0.638

Standard errors clustered at the store level in parentheses.

All interaction variables are mean-centered.

***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 8: Treatment effects on Social Cohesion, Helping, Peer pressure, and Job Satisfaction

	Social Cohesion	Helping	Peer Pressure	Job Satisfaction
Panel A	(1)	(2)	(3)	(4)
Treatment	0.040	0.233**	-0.124	0.013
	(0.091)	(0.104)	(0.139)	(0.096)
Employee-fixed effects	yes	yes	yes	yes
Questionnaire-fixed effects	yes	yes	yes	yes
Observations	570	564	512	554
Employees	285	282	256	277
R^2	0.003	0.026	0.003	0.005
Panel B	(5)	(6)	(7)	(8)
Treatment	0.051	0.236**	-0.125	0.019
	(0.090)	(0.104)	(0.139)	(0.094)
Treatment X Social cohesion pre-quest.	-0.034	-0.063	-0.014	-0.052
	(0.075)	(0.090)	(0.104)	(0.086)
Post-quest. X Social cohesion pre-quest.	-0.236***	-0.025	0.007	-0.089
	(0.052)	(0.076)	(0.037)	(0.072)
Employee-fixed effects	yes	yes	yes	yes
Questionnaire-fixed effects	yes	yes	yes	yes
Observations	570	564	512	554
Employees	285	282	256	277
R^2	0.102	0.034	0.003	0.013

Standard errors clustered at the store level in parentheses.

Post-quest. is a dummy variable indicating that an observation stems from the post-experiment questionnaire. Social cohesion pre-quest. is social cohesion at the individual level as measured in the pre-experiment questionnaire.

All interaction variables are mean-centered.

***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table 9: Robustness checks

	Placebo		Before-after		Cluster at group level		Social cohesion excl. managers	Correcting for dispersion in social cohesion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Dependent variable: sales growth				
Treatment	-0.717	-0.555	1.520	1.650	1.495	1.628	1.676	1.560
	(1.930)	(1.942)	(1.216)	(1.212)	(1.299)	(1.279)	(1.243)	(1.225)
Treatment X Social Cohesion		1.372		2.782		2.891*	2.817*	2.650
		(2.29)		(1.682)		(1.730)	(1.679)	(1.941)
Exp. period X Social Cohesion		-2.228		-2.252*		-2.365*	-1,877	-2.666*
		(1.763)		(1.214)		(1.284)	(1.282)	(1.472)
Treatment X st.dev Social Cohesion								1.187
								(3.527)
Exp. period X st.dev Social Cohesion								0.976
								(2.174)
Store-fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Week-fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Store-week observations	4990	4908	244	240	5722	5628	5534	5346
Stores	122	120	122	120	122	120	118	114
R^2	0.646	0.645	0.513	0.523	0.638	0.638	0.635	0.634

Standard errors clustered at the store level in parentheses in all columns except (5) and (6). In columns (5) and (6), standard errors are clustered at the competition group level.

All interaction variables are mean-centered.

***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Table A1: Regression of stores' social cohesion on other store characteristics

	Dependent variable: social cohesion	
	(1)	
Task cohesion	0.263**	(0.104)
Directive leadership	-0.091	(0.109)
Supportive leadership	0.229***	(0.080)
Number of employees	0.048	(0.029)
Average tenure employees	0.084	(0.055)
Std tenure employees	-0.046	(0.048)
Average age employees	-0.022	(0.029)
Std age employees	-0.019	(0.034)
Fraction male employees	0.448	(0.869)
Std gender of employees	-0.515	(0.791)
Fraction on-call employees	-0.376	(0.677)
Tenure manager	0.022*	(0.011)
Age manager	-0.008	(0.010)
Gender manager (male =1)	-0.081	(0.183)
Stores	116	
R^2	0.288	

Standard errors clustered at the store level in parentheses.

***, **, * denote statistically significant effects at the 1%, 5%, and 10% level, respectively.

Appendix

Survey questions (translated from Dutch)

Welcome to the survey for employees of <company>, executed by the University of Amsterdam.

We ask you to complete the questionnaire below attentively. The questionnaire is strictly confidential. Your answers as well as store averages will remain confidential and will not be shared with <company>.

Please indicate for each proposition which answer corresponds best to your situation in the past two months.

The following questions revolve around the effort in your store.
<answers: 7-point scale, completely agree … completely disagree>

- Our team is united in trying to reach its goals.
- I am not satisfied with my team's level of commitment to the tasks.
- Our team members have conflicting aspirations for the team's performance.
- This team does not give me enough opportunities to improve my performance.

The following questions revolve around the atmosphere in your store.
<answers: 7-point scale, completely agree … completely disagree>

- Our team would like to spend time together outside of work hours.
- Members of our team do not stick together outside of work hours.
- Our team members rarely do something nice together.
- Members of our team would rather go out on their own than get together as a team.
- For me this team is one of the most important social groups to which I belong.
- Some of my best friends are in this team.

<next page>

The following questions revolve around situations in which a co-worker does not work as hard as he/she should.
<answers: 7-point scale, completely agree … completely disagree>

- It is easy to see whether co-workers are working hard.

What did you do in the past two months in situations where a co-worker did not work as hard as he/she should? Please indicate to which extent the following propositions capture your actions in these situations.
<answers: 7-point scale, completely agree … completely disagree, and 'not applicable' >

- I did nothing.
- I talked directly to the co-worker.
- I spoke about it with the store manager.
- I spoke about it with other co-workers.

The following questions revolve around the helpfulness of employees in your store.
<answers: 7-point scale, completely agree … completely disagree>

- Team members go out of their way to help co-workers with work-related problems.
- Team members voluntarily help new employees settle into the job.

The following questions revolve around the role of the store manager.
<answers: 7-point scale, completely agree … completely disagree>

- The store manager expects his/her employees to follow instructions precisely.
- The store manager gives employees no room to make decisions on their own.
- The store manager supervises employees very closely.
- The store manager works hard to ease tensions in the team.
- The store manager works to develop close personal relationships with all employees.
- The store manager demonstrates concern for employees.

The following question revolves around your satisfaction with your job.

- How satisfied are you with your current job.
  <answers: 7-point scale, very satisfied …. very dissatisfied>

- Did you look for another job outside <company> in the past two months?
  <answers: yes, no>

- How likely is it that you still work at <company> next year?
  <answers: 7-point scale, very likely …. very unlikely>

Thank you for your time to participate in this survey.

Your answers have been recorded.

<The survey for store managers was almost identical. In the part on co-worker shirking, we removed the question 'I spoke about it with the store manager'. In the part on the role of the store manager, we replaced 'The store manager' by 'I' in all 6 questions.>