DISCUSSION PAPER SERIES

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

Josse Delfgaauw
Robert Dur
Oke Onemu
Joeri Sol

DISCUSSION PAPER SERIES

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

Josse Delfgaauw
Erasmus University Rotterdam and Tinbergen Institute

Robert Dur
Erasmus University Rotterdam, Tinbergen Institute, CESifo and IZA

Oke Onemu
Leiden University

Joeri Sol
University of Amsterdam and Tinbergen Institute

JULY 2020

ABSTRACT

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

We conduct a field experiment in a Dutch retail chain of 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 increases with the team's social cohesion, because social cohesion reduces free-riding behavior. In addition, 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 increases with social cohesion as measured before the intervention. Social cohesion itself is not affected by the team incentives.

Corresponding author:
Robert Dur
Erasmus University Rotterdam
Department of Economics H9-15
P.O. Box 1738
3000 DR Rotterdam
The Netherlands

E-mail: dur@ese.eur.nl

1   Introduction

Teamwork is of vital importance in many organizations, generating synergies in production, mutual learning, and social interactions with co-workers. But if 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 employers and employees.[1]

In this paper, we combine field experimentation and collection of rich questionnaire data to study two key questions: How does social cohesion among co-workers affect the effect of team incentives on team performance? And how do 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. If workers care more about their co-workers, they internalize the benefits of their efforts that accrue to co-workers to a larger extent. Hence, social cohesion reduces free-riding, implying that the effect of team incentives on performance is larger in more cohesive teams.

We test this theoretical prediction by conducting a field experiment in a large retail chain in The Netherlands comprising 122 stores. We introduce short-term team incentives in a randomly selected subset of these stores. To filter out common shocks, the team incentive is designed as a contest in 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.[2] 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 with the social cohesion of the team. In the analysis, we control for a rich set of possibly confounding factors, i.e. for store and team characteristics that may 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

---

[1]Employers increasingly acknowledge the importance of social relations in the workplace. Many organizations 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 tend to value good social relations with co-workers. Various studies have shown that 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, Krueger and Schkade 2008, Deversi et al. 2020).

[2]More precisely, the performance metric in the contest is the percentage growth in sales, where the reference period is the same weeks in the previous calendar year. We use this metric because it is the main store-level performance metric used in this retail chain.

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 coerce their co-workers into exerting higher effort (Kandel and Lazear 1992, Barron and Gjerde 1997, Freeman et al. 2010, Carpenter et al. 2018). Even if peer pressure is effective in raising team performance, social cohesion may suffer, in particular when peer pressure works through engendering feelings of shame and guilt. Alternatively, team incentives may induce team members to help each other at work (FitzRoy and Kraft 1986, Drago and Turnbull 1988, Harrison et al. 2002, Berger et al. 2011) and to invest (more) in co-worker social relations (Rotemberg 1994, Dur and Sol 2010).[3] 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.[4]

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). The 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).[5] 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. (1985), 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.[6]

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, but it

---

[3] The underlying idea behind this prediction from Rotemberg (1994) and Dur and Sol (2010) is that workers realize that they suffer from each other's free-riding behavior in response to the team incentive, and that one way to reduce free-riding behavior is to increase co-worker altruism. Workers may engage in several behaviors to try to increase co-worker altruism. They may show more interest in each other, they may help each other more, or they may spend more time together outside work hours. Obviously, none of these behaviors ensure that co-worker altruism will increase, but they may make it more likely.

[4] While we did not preregister our study (this was relatively uncommon when we designed our experiment in 2013), it will be clear from the design of the experiment (in particular, the stratified randomization and the content of the questionnaires) that our hypotheses were formulated before running the field experiment, not after.

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

[6] 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 also 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.

falls well within the range of estimated treatment effects in earlier studies in similar contexts.[7]

Second, as predicted, the effect of the team incentive on sales increases with the team's pre-existing level of social cohesion. This association is sizable: a one standard deviation increase in social cohesion is associated with a two percentage points higher estimated effect of the team incentives on sales growth. 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 coincide with weaker free-rider effects.

Third, 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 evidence for a small increase in co-worker helping, while peer pressure and job satisfaction are unaffected. Hence, social relations among team members are neither improved nor hurt by the implementation of team incentives.

Our paper contributes to several 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) finds evidence for substantial free-riding when individual incentives are replaced by team incentives for orange pickers. Further, Erev et al. (1993) and Bandiera et al. (2013) show 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 (Delfgaauw et al. 2013, 2014, 2015, 2020; 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, which studies how social concerns affect workers' performance (see Ashraf and Bandiera 2018 for a recent

---

[7]See Delfgaauw et al. (2013, 2014, 2015, 2020) and Friebel et al. (2017). Using levels of team incentive pay that are quite comparable to those used in our experiment, average treatment effects on sales performance range from 0% to 5%. A power analysis shows that we can detect an effect size of 3 percentage points with power 0.8, see Section 5.1.

review). An influential 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, suggesting that workers derive nonpecuniary gains from team production. 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.[8] Babcock et al. (2015) studies how students can be motivated to visit the gym and the library more often and find that team incentives work better than individual incentives when team members know each other, while the opposite holds when the team mate's identity was unknown (see also De Paola et al. 2019). 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. Like these studies, our field experiment shows that the effect of incentives on performance crucially depends on the social relations among team members.[9] 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 in Japan. They find that when workers share output (i.e. pool their catch at the end of the day), they tend to have stronger pro-social preferences. Moreover, they find evidence supporting the prediction that pro-social preferences among team members foster team output. Given that many employers and employees strongly care about social relations at work (see footnote 1 above), it is surprising how little causal evidence exists on how organizational policies affect social relations among employees.[10]

---

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

[9]Our study also relates to the literature on group identity and in-group favoritism (Tajfel and Turner 1979, Akerlof and Kranton 2000; see Charness and Chen 2020 for a recent survey). 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.

[10]Since we have no exogenous variation in social cohesion across teams in our study, we cannot assess whether strengthening social cohesion leads to a change in team performance. Meta-analyses of the literature in organizational psychology and management distill a positive relationship, but acknowledge mixed results of individual studies (Mullen and Copper 1994, Gully et al. 1995, Beal et al. 2003, Bakundi and Harrison 2006, Bell 2007, Chiocchio and Essiembre 2009). More recently, a number of lab experiments have been

Lastly, our paper also contributes to the literature on complementarities of human resource management practices. Englmaier and Schüßler (2016) review this nascent literature and suggest that behavioral economics is well-suited to explore how such complementarities can explain persistent productivity differences within industries. A recent example is the field experiment by Blader et al. (2020), who show that the effects of publicly posting individual employee's performance crucially depend on the local "collectivist orientation." More generally, Blader et al. (2020) argue that the success of management practices is contingent on the prevailing company culture. Likewise, our field experiment reveals important complementarities between team incentives and a team's social cohesion.

For managers, our findings suggest that team incentives work better when social cohesion in a team is higher. Moreover, managers could consider to foster social cohesion among workers in order to reap more rewards of team incentives. Our result that team incentives neither improved nor harmed social cohesion within stores 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.

The remainder of the paper is structured 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 of main findings and implications.

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.[11] All stores of the retail chain are located in The Netherlands and operate under the same brand name. The stores sell clothes, shoes, and items for sports and outdoor activities. The chain mostly targets budget-minded consumers. 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.

---

performed studying how team performance depends on the opportunities for peer communication. 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. Similarly, Corgnet et al. (2020) finds that endowing teams with peer chat reduces shirking and increases production.

[11]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.

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. 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, advising customers, and cleaning the store. In our context, 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 (including the store managers) 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 the randomization by stores' social cohesion (as measured in the pre-experiment survey) and by stores' past sales performance. We explain the randomization procedure in detail 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 in the previous calendar year. We chose this performance metric because it is well-known in this retail chain among both the store managers and the employees. The bonus was set at 75 euro for full-time employees, 50 euro for part-time employees, and 25 euro for on-call employees.[12] For full-time employees,

---

[12]Employees would receive this bonus independent of how many hours they actually worked during the tournament period, implying that (self-)selection effects are likely limited.

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 such as Delfgaauw et al. (2013) and Friebel et al. (2017); see the Discussion section in Delfgaauw et al. (2020) for a detailed overview. The treatment period was week 42 to 47 in 2013 (October 14th to November 24th). During this period, no other major 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 national holidays, the weather, and advertisement campaigns of this company and those of competing companies. Figure 1 shows the weekly average of all stores' sales growth for the period of week 1 to week 47 in 2013 and its variation across stores within weeks.[13] 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 pre-determined absolute sales targets. A positive (negative) shock can make a pre-determined target too easy (difficult) to reach, weakening 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).

Theory predicts that the effects of tournament incentives are stronger if contestants are more homogenous in ability (Lazear and Rosen 1981). Therefore, we assigned treatment stores to groups based on their historical performance in the following way. We ranked the treatment 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 the 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 assigned to the same group. Thus, seven stores were re-assigned.

All communication on the experiment to employees ran through the company's regular communication channels. One and a half weeks before the start of the contests, a general announcement was sent digitally to all employees. The message, which was signed by the HR-manager, stated that several incentive events would take place in the coming period, that only a subset of stores would participate in each event, and that more information would follow soon. Four days before the start, all store managers received a (hardcopy) letter, signed by the commercial manager and the HR-manager of the retail chain. The letter for managers of stores in the control group stated that their store was randomly selected to

---

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

not take part in the incentive event this time, but that their store would take part in the next event.[14] 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. The Online Appendix provides the exact texts of all letters (translated from Dutch).

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 name and 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. Figure 2 provides an example of these posters.[15] Store managers were instructed to hang the posters at 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.[16] As sales growth is a well-known and widely used performance measure in this 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 among workers within a store. Furthermore, we are interested in whether the introduction of team incentives results in a change in co-worker helping and/or peer pressure. For this reason, we administer a pre-experiment and

---

[14]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. Regarding possible ratchet effects, we cannot completely rule them out, but consider their existence not very likely in our context. Our performance measure is percentage growth in sales as compared to sales in the same weeks one year ago. Hence, when a team would attempt to manipulate its baseline performance so as to make it easier to achieve higher sales growth in a future tournament, it would benefit only with a lag of a full year and only if it would take that long before it was the team's turn to take part in an incentive event. The digital communication stated that the incentive events would take place "in the coming period". We consider it not very likely employees may have believed it would take more than a year for them to take part. Finally, the performance measure was not communicated to control stores. Hence, they would have had to guess what performance to manipulate (even though sales growth would probably be seen as a likely measure).

[15]The negative sales growth numbers shown in Figure 2 are in line with the relatively weak sales growth in these stores in the entire calendar year. Furthermore, Figure 1 shows that the experimental period was a weak period overall, albeit not extraordinarily so.

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

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 Online Appendix.

The literature offers several survey measures for social cohesion in work teams (see Casey-Campbell and Martens 2009 for a review). We use the survey questions developed by Carless and De Paola (2000), as listed in Table 1. The questions measure friendship among workers and whether workers like to interact with each other outside of work hours. We take the average of the responses to the six questions as an individual's perception of his or her team's social cohesion.[17]

Social cohesion in work teams 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 therefore include 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. Supportive 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. To this end, we take the mean of the individuals' average score on a given survey measure over all respondents from a given store. As employees may perceive the level of social cohesion in their store differently, in a robustness check we control for the variance in social cohesion as expressed by employees from the same store.[18] 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).

---

[17] In taking averages 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.

[18] In the main analysis, we make no distinction between managers and employees in determining teams' social cohesion and other store-level survey outcomes. As a further robustness check, we also construct a measure of social cohesion that excludes the managers' responses.

Their 'anti-shirking index' measures the willingness to undertake action upon observing a shirking colleague, using 4 questions.[19] For co-worker helping, we use 2 out of 5 questions from the scale for interpersonal helping developed by Moorman and Blakely (1995), who used this scale as one dimension in their measurement of organizational citizenship behavior. 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. 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).

We sent e-mail invitations to take part in the pre- and post-experiment online questionnaires to all employees and store managers. The e-mail stated that the company collaborated with the University of Amsterdam to investigate the satisfaction of all employees in the stores. The complete e-mails can be found in the Online Appendix. We sent two reminders to non-respondents. For 16 percent of the employees, we did not have an email-address. These employees received invitations to complete the online questionnaire through regular mail. The invitation letter was identical to the e-mail, apart from a personalized password to access the online questionnaire. We did not send reminders through regular mail.[20]

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. To this end, 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.[21] We created 12 blocks of stores with similar social cohesion, i.e. 10 blocks with 10 stores and 2 blocks with

---

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

[20]It is unlikely that employees linked the questionnaires to the incentive event, as the company has organized similar sales competitions in the past. Neither the questionnaires nor the accompanying emails mentioned the sales competitions or team incentives. Moreover, the pre-experiment survey was conducted before any details on the incentive event were communicated. Hence, we maintain that our experiment can be classified as a natural field experiment (Harrison and List 2004). If some employees, particularly in the treatment group, did link the post-experiment survey to the incentive event, this could affect their decision to participate in the survey. However, we observe no differential self-selection into the surveys between the treatment and control group, as discussed in section 3.2 below.

[21]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.

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 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 percentage sales growth compared to the same week a year earlier.[22] 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 on sales 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, the company in our study 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 on average 27 years old and have 6 years of tenure. About 40 percent of

---

[22]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 earlier is calculated as $g_{s,w} = \frac{I_{s,w,2013} - I_{s,w,2012}}{I_{s,w,2012}} \times 100 = \frac{\frac{R_{s,w,2013}}{R_{s,b}} - \frac{R_{s,w,2012}}{R_{s,b}}}{\frac{R_{s,w,2012}}{R_{s,b}}} \times 100 = \frac{R_{s,w,2013} - R_{s,w,2012}}{R_{s,w,2012}} \times 100$, which is independent of base week sales $R_{s,b}$.

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.[23]

3.2 Survey measures

Panel A and panel B of Table 3 show the descriptive statistics of the pre- and post-experiment survey, respectively. Column 1 shows that, across stores, on average 35 percent of the employees in a store completed the first survey. About four out of five managers completed the first survey. The second survey is completed by, on average, 23 percent of employees and 68 percent of managers.[24] 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 appear, on average, similar in historical performance, personnel characteristics, and team characteristics as measured by our survey. Panel B of Table 3 shows that in the post-experiment survey, the outcomes are quite close to the pre-experiment outcomes.[25]

Table A1 in the Appendix provides correlations between store performance, personnel characteristics, and the survey measures. Average sales growth before the experiment is

---

[23]Figure A1 in the Appendix shows a common trend in sales growth for treated and control stores in the period before the experiment.

[24]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.

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

neither correlated with social cohesion nor with task cohesion, but is weakly positively (negatively) correlated with supportive (directive) leadership.[26] Social cohesion correlates positively with task cohesion, supportive leadership, and co-worker helping and negatively with directive leadership, peer pressure, and the standard deviation in employees' age. Generally, personnel characteristics neither correlate strongly with sales growth nor with the survey measures.

3.2.1 Validity of survey measures

Our survey measures have been tried-and-tested in earlier literature. In Table A2 in the Appendix, we report three internal validity statistics. First, we determine Cronbach's Alpha for all measures that consist of more than one question. 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. Second, we report intra-class correlation coefficients, which provide an indication of the strength of within-store correlations in the survey measures. The intra-class correlation of social cohesion is around 0.2. Hence, within a team, employees' perception of the social cohesion in the team contains a common element, but also contains idiosyncratic elements. For most other survey measures, the intra-class correlation is rather small as well.

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, using the data from the control group only.[27] 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 results are qualitatively affected if we drop directive leadership from our analysis.

3.2.2 Self-selection into the surveys

Given our design, employees' self-selection into the pre- and post-experiment surveys raises several possible concerns. 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

---

[26]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.

[27]We do not use the data from the treatment group here because the post-experiment measures may be affected by the treatment.

before the stores' assignment. However, 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. We cannot fully rule out a bias due to self-selection, as it may be driven by unobservables. However, the findings discussed below show that on all observable dimensions, self-selection into the surveys appears unrelated to assignment to treatment or control.

First, Table A3 in the Appendix shows that there are 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, see Table 3.

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.[28] Conditional on having participated in the pre-experiment (post-experiment) survey, the probability of responding to the post-experiment (pre-experiment) survey is 62% (87%) in the treatment group and 57% (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.[29] Hence, self-selection into the post-experiment (pre-experiment) survey appears unrelated to employees' responses to the pre-experiment (post-experiment) survey.

3.3 Balance test and pre-trends by level of social cohesion

For identifying the relation between social cohesion and stores' response to team incentives, it is important that along the social cohesion dimension treatment and control stores are similar in all other respects. In Sections 3.1 and 3.2, we showed that stores in the treatment group and the control group were on average comparable in all observable characteristics and in sales leading up to the experimental period. Table A4 in the Appendix gives descriptive statistics by treatment status separately for stores with high and low pre-experiment social cohesion. Among all personnel characteristics and pre-experiment survey measures, we find

---

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

[29]Performing these tests separately for respondents from stores in the treatment group and respondents from stores in the control group does not yield any statistically significant difference either.

one statistically significant difference between treatment and control stores: For high-social cohesion stores, pre-experiment job satisfaction is significantly higher among treatment stores than among control stores. To exclude that our findings are driven by this difference, in a robustness check in Section 6 we control for the interaction between the treatment and pre-experiment job satisfaction.

Figure A2 in the Appendix gives average weekly sales by treatment status for each social cohesion quintile. Panel A covers the full period, showing no sign of differential pre-trends between treatment and control stores in any of the quintiles.[30] Panel B zooms in on the final 12 weeks of our sample period, where the lower degree of sales volatility allows for a better view of the common trends leading up to the experimental period.

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 by employees. We estimate the effect of team incentives on sales growth using panel OLS regressions with store- and week-fixed effects. Let $g_{s,w}$ denote the percentage sales growth of store s in week w. The average treatment effect is estimated by:

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

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 for teams with higher social cohesion. As discussed in the Introduction, in more cohesive teams we expect a stronger response to team incentives as workers internalize the benefits of their effort that accrue to their co-workers to a larger extent (Rotemberg 1994; Dur and Sol 2010).

Let C_s^q 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 social cohesion and the effect of team incentives on performance by interacting the

---

[30] Figure A2 does show that in the third quintile, average sales growth in the treatment group lies consistently above average sales growth in the control group. This is due to one store in the treatment group that has an average sales growth of more than 50 percent (59 percent before the experiment and 54 percent during the experiment). Our results are qualitatively similar if we remove this store from the analysis.

treatment dummy $T_{s,w}$ with social cohesion as measured in the pre-experiment questionnaire C_s^1:

$$g_{s,w} = \alpha_s + \gamma_w + \beta T_{s,w} + \psi T_{s,w} C_s^1 + \phi E_{s,w} C_s^1 + \varepsilon_{s,w} \tag{2}$$

In equation (2), $E_{s,w}$ 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 treatment. 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 regressions that allow for a nonlinear relation between social cohesion and the effect of team incentives on performance. To facilitate the interpretation of the coefficients, we mean-center all continuous interaction variables.

Stores' social cohesion C_s^1 is taken from the pre-experiment survey and, therefore, is not affected by the introduction of team incentives. However, if C_s^1 is correlated with other store characteristics which may affect the response to team incentives, our estimate ψ is biased. To deal with such 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. These include task cohesion and leadership style, which have been shown in earlier literature to correlate strongly with social cohesion. Similarly, we control for interactions with store characteristics from the administrative data that may affect the response to team incentives, such as the number of employees in the team, the gender composition, and employees' tenure.[31]

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 team incentives can lead to more peer pressure. 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.

---

[31] 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.

We estimate the effects of team incentives on social cohesion, co-worker helping, peer pressure, and job satisfaction at the employee-level.[32] Let $C_{s,i}^q$ denote the social cohesion reported by 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. As before, we cluster standard errors at the store level. We estimate the average treatment effect on the other survey measures in the same way.

5 Results

5.1 Average treatment effect

Column 1 in Table 4 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 it is well within the range of findings in comparable settings (see for an overview the Discussion section in Delfgaauw et al. 2020). A power analysis shows that we can detect an effect size of 3 percentage points with power 0.8.[33]

5.2 How does social cohesion relate to the effect of team incentives on performance?

Before presenting the estimates of equation (2), Figure 7 provides a graphical presentation of the heterogeneity in the treatment effect. For each store, we calculate the difference between average sales growth during the experimental period and average sales growth during all weeks before the experimental period. Figure 7 plots this difference in sales growth against the stores' pre-experiment social cohesion for treated stores (dots) and control stores (squares). We find that for social cohesion below 4 there appears to be little systematic

---

[32] We also discuss the results of performing these estimations at the store level.

[33] 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 (see e.g. List et al. 2011), this yields a minimal detectable effect size of 2.96 percentage points.

difference in sales performance between treatment and control stores. However, for social cohesion above 4, treated stores tend to show a bigger improvement (or smaller worsening) of sales performance during the experimental period as compared to similarly cohesive stores in the control group.

Column 2 in Table 4 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 with 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 (0.71 on a 7-point scale) corresponds to an increase in the predicted effect of team incentives on sales by 2.1 percentage points.[34]

Column 3 of Table 4 adds interaction effects for task cohesion and leadership to the estimation in column 2. The estimated interaction effect of team incentives and social cohesion is hardly affected. Furthermore, task cohesion and employee-manager relations appear to be not very predictive of 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

---

[34]Note that the estimation results also show that, unconditional on treatment, sales growth during the experimental period is lower in stores with higher social cohesion (see the significant and sizeable coefficient for the interaction between experimental period and social cohesion in the second column of Table 4). The placebo test reported in Section 6 below shows that this pattern in the data already existed at least six weeks before the experimental period, ruling out that contamination drives this relation. We can think of several possible reasons for why sales growth in stores with higher social cohesion is relatively low (or why sales growth in stores with lower social cohesion is relatively high) during this period. One reason could be the worsening labor market conditions. Unemployment was on the rise in the Netherlands during our sample period. As a result, employees may have become more scared to lose their job and, consequently, chose to shirk less and work more (Shapiro and Stiglitz 1984). This may particularly hold for employees in stores with low social cohesion, because they may attach little special value to keeping the job during a tight labor market. In contrast, employees in stores with high social cohesion may care a lot about keeping their job also in a tight labor market, as they appreciate the social relationships at work (see also the empirical evidence in footnote 1) and cannot be sure to find a similar work environment in a new job. As a result, in stores with high social cohesion, employees worked hard already during a tight labor market and so may have had less opportunities to improve performance when the labor market situation worsened as compared to stores with low social cohesion. Note that this reasoning (which relies on a nonlinearity in the performance-effort relationship) would also imply that the effect of team incentives depends less strongly on a team's social cohesion than it otherwise would (when the performance-effort relation would be linear). A second possible reason for why sales growth is relatively low in stores with high social cohesion during this period is the weather. There was a lot of rain during the Fall of 2013 in the Netherlands. This may have affected the number and type of customers visiting the stores. Presumably, on rainy days, less customers visit the stores and those who do may be more dedicated to buy at this store. Sales performance on such days may be less dependent on the team's social cohesion than on days without rain, as the inflow of more customers who are less dedicated to buy may require more cooperation and team spirit to achieve good sales performance. For the same reason, we may have seen an even stronger relation between social cohesion and the effect of team incentives on sales performance if the weather during the experimental period had been better, as it would have provided more opportunities for stores with high social cohesion to excel.

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 larger. None of the store characteristics relates significantly to the response to team incentives. While we obviously cannot completely rule out any remaining omitted variable bias, these findings raise our confidence that such possible bias is limited.[35]

The estimations in columns 2 to 4 in Table 4 assume a linear relation between social cohesion and the response to team incentives. In Figure 8, 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.[36] Clearly, 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.

Panel A of Figure A3 in the Appendix gives the results of another way to estimate a non-linear relation between social cohesion and the response to team incentives: estimating the average treatment effect within each social cohesion quintile. As the number of stores per quintile is only 24, the power of these estimations is rather low and none of the estimated effects are statistically significant. Still, the pattern is similar to that in Figure 8: The treatment effects are close to zero in quintiles 1 to 4 and markedly higher for the top quintile.[37]

5.3   Do team incentives affect social cohesion?

Panel A in Table 5 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 (at the individual level). 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

---

[35] As a final robustness check, we regressed social cohesion on task cohesion, leadership styles, and all store characteristics in our data, see Table A5 in the Appendix. Next, we interacted the residuals of this regression with the treatment dummy and the experimental period 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 4: a coefficient of 3.17 with a standard error of 1.91.

[36] The exact specification is $g_{s,w} = \alpha_s + \gamma_w + \beta T_{s,w} + \psi T_{s,w} C_s^1 + \chi T_{s,w} \left(C_s^1 \right)^2 + \phi E_{s,w} C_s^1 + \xi E_{s,w} \left(C_s^1 \right)^2 + \varepsilon_{s,w}$.

[37] A downside is that the estimated effects are sensitive to the exact binning of stores. For instance, if we would use quartiles rather than quintiles, the estimated treatment effect for the fourth quartile would be statistically significant at the 10%-level. Yet, the average treatment effect of the second quartile would be markedly higher than the estimated effect for the first and third quartile. Figure 7 shows that such non-monotonicity is not actually present in the data.

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.[38]

We find similar results if we exclude the store managers from these estimations. 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.[39]

Panel B in Table 5 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.[40] Again, using store-level measures rather than individual-level measures does not affect these findings qualitatively.

Overall, the effects of the short-term 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. Whether these results continue to hold if team incentives are introduced for a longer period, or even permanently, is an important question for further research.

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 6 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 4,

---

[38]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. The positive effect on helping remains statistically significant (as well as economically small) after applying the Bonferroni correction for multiple hypotheses testing.

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

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

columns 1 and 2) and not statistically significant.[41] Panel B of Figure A3 in the Appendix gives the average placebo-treatment effects by social cohesion quintile. For the first quintile, the placebo-effect is similar to the actual treatment effect presented in Panel A. In the other quintiles, the placebo-effects are (much) smaller than the actual treatment effects. Most notably, the estimated placebo-treatment effect for the highest social cohesion quintile is close to zero.

Second, in column 3 and 4 of Table 6 we drop store-fixed effects from our estimations and include a dummy for stores in the treatment group, stores' pre-experiment social cohesion, and the interaction between social cohesion and the treatment group dummy. The results show that this neither affects the estimated average treatment effect of the introduction of team incentives nor the estimated relation between social cohesion and the treatment effect.[42]

Third, we create a measure of stores' social cohesion that excludes 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 5 of Table 6.

Fourth, 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 differently to team incentives. Lacking individual measures of performance, we cannot directly assess this. Instead, we analyze whether controlling for the interaction with the pre-experiment within-store dispersion in reported social cohesion affects our findings. We use the within-store standard deviation of social cohesion at the individual level as the measure for dispersion.[43] Column 6 of Table 6 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.

---

[41] Note also that the coefficient for the interaction between the experimental period and social cohesion in column 2 of Table 6 has approximately 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 common pre-existing trend, ruling out that contamination drives this relation.

[42] Using the same approach for the estimations reported in columns 3 and 4 of Table 4 also yields similar results. Note also that the interaction between the treatment store dummy and social cohesion is small and statistically insignificant, implying that in the period before the introduction of team incentives the relation between stores' social cohesion and sales growth did not differ between the treatment group and the control group.

[43] 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.

The within-store dispersion in social cohesion itself does not have a statistically significant relation with the response to team incentive.

Table A6 in the Appendix shows the results of four more robustness checks. First, 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). Second, we cluster standard errors at the competition group level for treatment stores, to correct for possible correlation in error terms during the experimental competition.[44] Third, we winsorize sales growth to account for possible outliers. Fourth, we control for the interaction between pre-experiment job satisfaction and the treatment effect, as we found that job satisfaction in stores with high social cohesion was significantly higher in the treatment group than in the control group (Table A4). For all robustness checks, the magnitudes of the estimated effects are of similar magnitude as in our main estimations, although the interaction between social cohesion and the treatment dummy just loses statistical significance at the 10 percent level in the before-after estimation and the winsorized estimation (p-values of 0.101 and 0.104, respectively).[45]

7   Concluding remarks

By combining rich questionnaire data and a large-scale field 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 in teams where team members have better social relations. This result is in line with the theoretical prediction that in teams with good social relations, individuals internalize 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 may benefit from enhancing social interactions among employees.

In interpreting our results, it is important to keep in mind that our field setting did not allow us to experimentally manipulate teams' social cohesion. As a result, we cannot completely rule out omitted variable bias, i.e. the possibility that social cohesion is correlated with some unobserved variable that is a determinant of the response to incentives as well,

---

[44]We do not attempt to estimate dynamic effects induced by intermediate rankings during the competition. As intermediate rankings 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. See also Casas-Arce and Martinez-Jerez (2009).

[45]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 8.

thus biasing our estimate. To reduce the scope for this possible bias, we collected data on what the literature suggests are important correlates of social cohesion, and found that our results are robust to inclusion of these variables interacted with the treatment. A challenge for future work is to find or generate exogenous variation in social cohesion in work teams prior to providing team incentives to a random subset of the teams. Babcock et al. (2015), De Paola et al. (2019), and Corgnet et al. (2019, 2020) have taken important steps in that direction.

The theory that inspired our field experiment predicts that altruism among workers strengthens the response to team incentives, because altruistic workers take into account that their work effort not only affects their own but also their co-worker's expected earnings. Alternative explanations for our supportive findings cannot be ruled out, however. For instance, altruistic workers may more easily coordinate their actions, because they tend to spend more time together. In our specific context, more cohesive teams may be more aware of the team incentive as they are more likely to jointly learn about the tournament when visiting the canteen together and agree upon being engaged in the tournament.

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.[46] More generally, given that both employers and employees value good social relations highly, learning how management practices affect social cohesion at work is an important area for future research.

---

[46]See, however, Carpenter and Seki (2011)'s study of fishermen in Japan which we discussed in the Introduction. Changes to incentive pay may also shape the composition of the workforce through signaling and self-selection, see e.g. Sliwka (2007) and Danilov and Sliwka (2017). A related literature studies the effects of workplace organization on social preferences, e.g. Gneezy et al. (2016).

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] Bakundi, Prasad, and David A. Harrison (2006), Ties, Leaders, And Time In Teams: Strong Inference About Network Structure's Effects On Team Viability And Performance, *Academy of Management Journal* 49(1): 49 – 68.

[5] 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.

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

[7] 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.

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

[9] 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.

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

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

[12] 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.

[13] Blader, Steven, Claudine Gartenberg, and Andrea Prat (2020), The Contingent Effect of Management Practices, *Review of Economic Studies* 87(2): 721–749.

[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] Carpenter, Jeffrey, Andrea Robbett, and Prottoy A. Akbar (2018), Profit Sharing and Peer Reporting, *Management Science* 64(9): 4261 – 4276.

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

[21] Casas-Arce, Pablo, and F. Asís Martínez-Jerez (2009), Relative Performance Compensation, Contests, and Dynamic Incentives, *Management Science* 55(8): 1306 – 1320.

[22] 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.

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

[24] Charness, Gary, and Yan Chen (2020), Social Identity, Group Behavior, and Teams, *Annual Review of Economics* 12: forthcoming.

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

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

[27] 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.

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

[29] Corgnet, Brice, Roberto Hernán-Gonzalez, and Ricardo Mateo (2019), Rac(g)e Against the Machine? Social Incentives When Humans Meet Robots, GATE working paper 2019-04.

[30] Corgnet, Brice, Brian Gunia, and Roberto Hernán González (2020), Harnessing the Power of Social Incentives to Curb Shirking in Teams, GATE working paper 2020-06.

[31] 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.

[32] Danilov, Anastasia, and Dirk Sliwka (2017), Can Contracts Signal Social Norms? Experimental Evidence. *Management Science* 63(2): 459 – 476.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] Delfgaauw, Josse, Robert Dur, and Michiel Souverijn (2020), Team Incentives, Task Assignment, and Performance: A Field Experiment, *The Leadership Quarterly* 31: 101241.

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

[39] 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.

[40] Deversi, Marvin, Martin G. Kocher, and Christiane Schwieren (2020), Cooperation in a Company: A Large-Scale Experiment, Mimeo, University of Munich (LMU).

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

[42] 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.

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

[44] 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.

[45] 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.

[46] Englmaier, Florian, and Katharina Schüßler (2016), Complementarities of Human-Resource Management Practices: A Case for a Behavioral-Economics Perspective, *Journal of Institutional and Theoretical Economics* 172(2): 312 – 341.

[47] 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.

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

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

[50] 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.

[51] 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.

[52] 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.

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

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

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

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

[57] Gneezy, Uri, Andreas Leibbrandt, and John A. List (2016), Ode to the Sea: Workplace Organizations and Norms of Cooperation, *Economic Journal* 126: 1856 – 1883.

[58] 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.

[59] 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.

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

[61] 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.

[62] 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.

[63] 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.

[64] 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.

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

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

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

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

[69] Krueger, Alan, and David Schkade (2008), Sorting in the Labor Market; Do Gregarious Workers Flock to Interactive Jobs?, *Journal of Human Resources* 43(4): 859 – 883.

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

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

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

[73] 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.

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

[75] 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.

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

[77] 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.

[78] 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.

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

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

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

[82] 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.

[83] 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.

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

[85] Shapiro, Carl, and Joseph E. Stiglitz (1984), Equilibrium Unemployment as a Worker Discipline Device, *American Economic Review* 74(3): 433 – 444.

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

[87] Sliwka, Dirk (2007), Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes, *American Economic Review* 97(3): 999 – 1012.

[88] 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.

[89] 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.

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

[91] Wendt, Hein, Martin C. Euwema, and Hetty van Emmerik (2009), Leadership and Team Cohesiveness across Cultures, *The Leadership Quarterly* 20(3): 358 – 370.

[92] 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.

[93] 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.

[94] 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.

[95] 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.

Figure 1: Average weekly sales growth

Average sales growth is the percentage change in weekly sales at the store level
compared to the same week in the previous calendar year, averaged over all stores.
The experiment took place in the period between the vertical red lines (weeks 42 to 47).
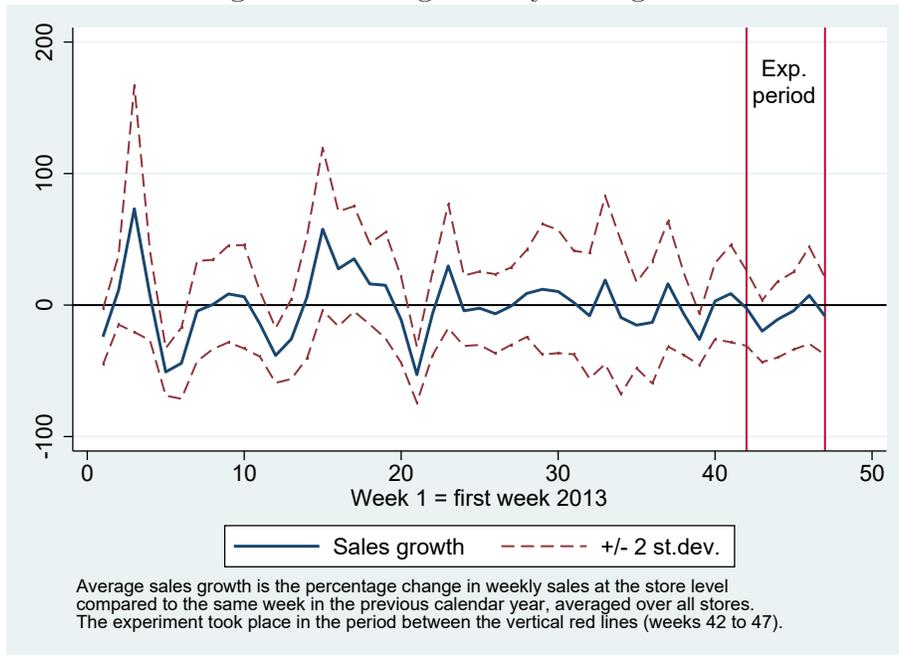
Figure 2: Example of the poster that treatment stores received every week during the contest
(translated)

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!**

Note: The numbers in brackets are store identifiers.

Figure 3: Design and timeline of the experiment

| Early Oct 2013: Pre-experiment questionnaire | Experimental period Oct - Nov 2013 | | Group 1: 4 stores | Early Dec 2013: Post-experiment questionnaire |

Treatment: 72 stores

Group 2: 4 stores

122 stores

Control: 50 stores

Group 18: 4 stores

Treatment assignment stratified by social cohesion and average sales growth over 6 months before experiment

Group assignment based on average sales growth over 6 months before experiment

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

Figure 4: Distribution of social cohesion across stores
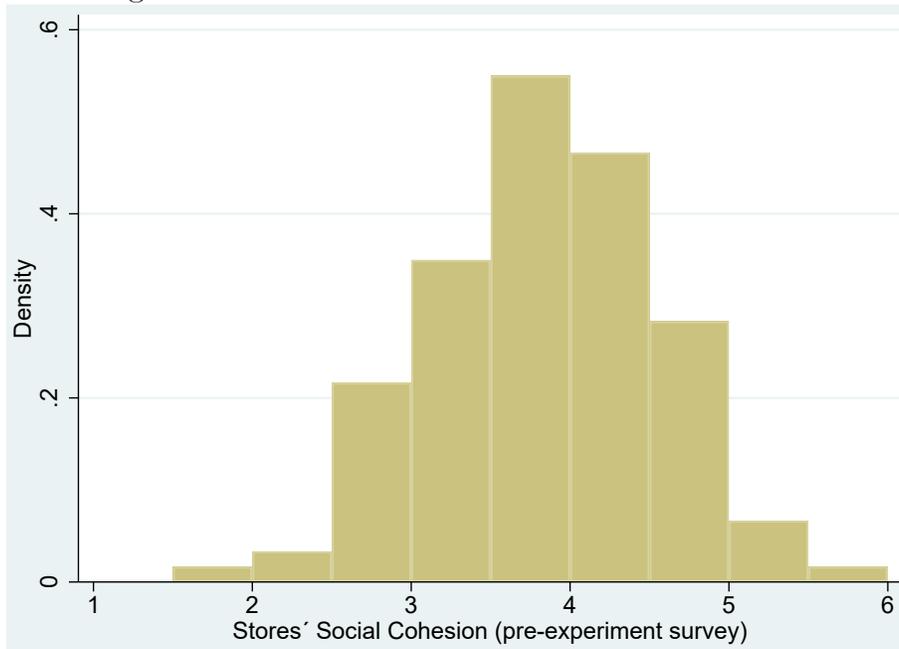
Figure 5: Kernel densities of pre-experiment survey measures by response to post-experiment survey

Solid lines: Respondents did not respond to post-experiment survey.
Dashed lines: Respondents did respond to post-experiment survey.
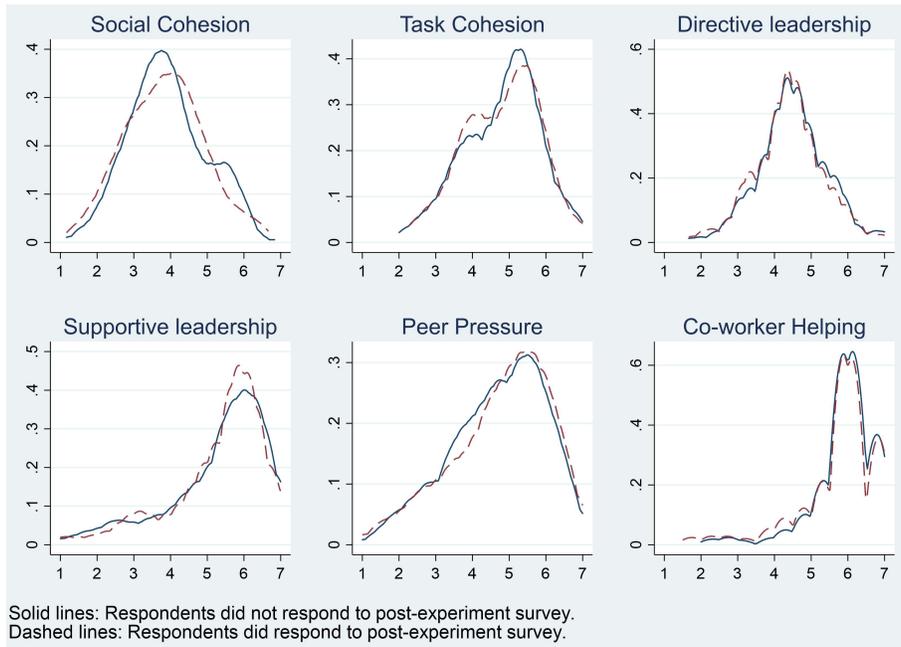
Figure 6: Kernel densities of post-experiment survey measures by response to pre-experiment survey

Solid lines: Respondents did not respond to pre-experiment survey.
Dashed lines: Respondents did respond to pre-experiment survey.
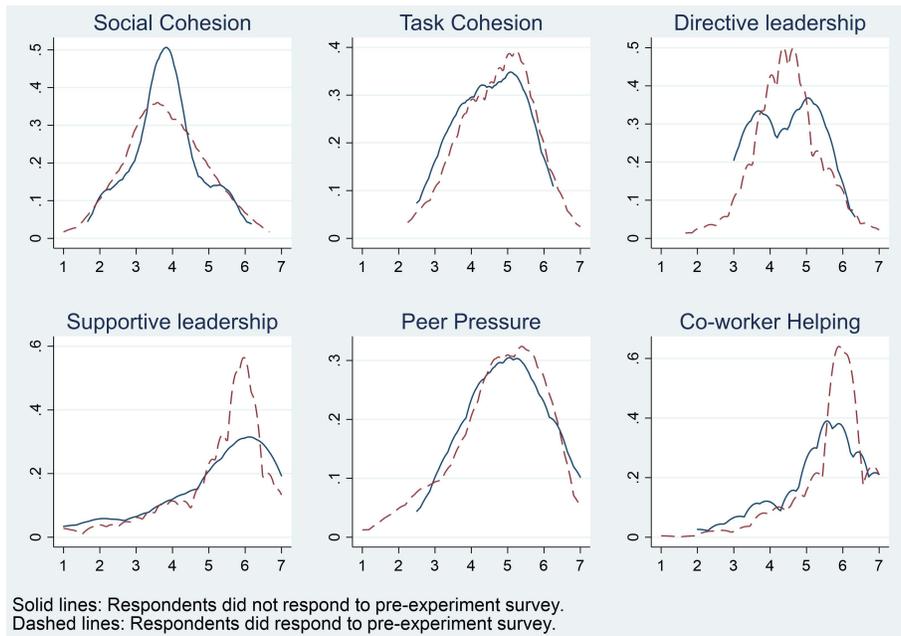
Figure 7: Scatterplot of stores' social cohesion and the difference between average sales growth during and before the experimental period, by treatment status

The x-axis gives a store's social cohesion as measured in the pre-experiment survey. The y-axis gives the difference between a store's average weekly sales growth during the experimental period and the average weekly sales growth over all weeks before the experimental period.
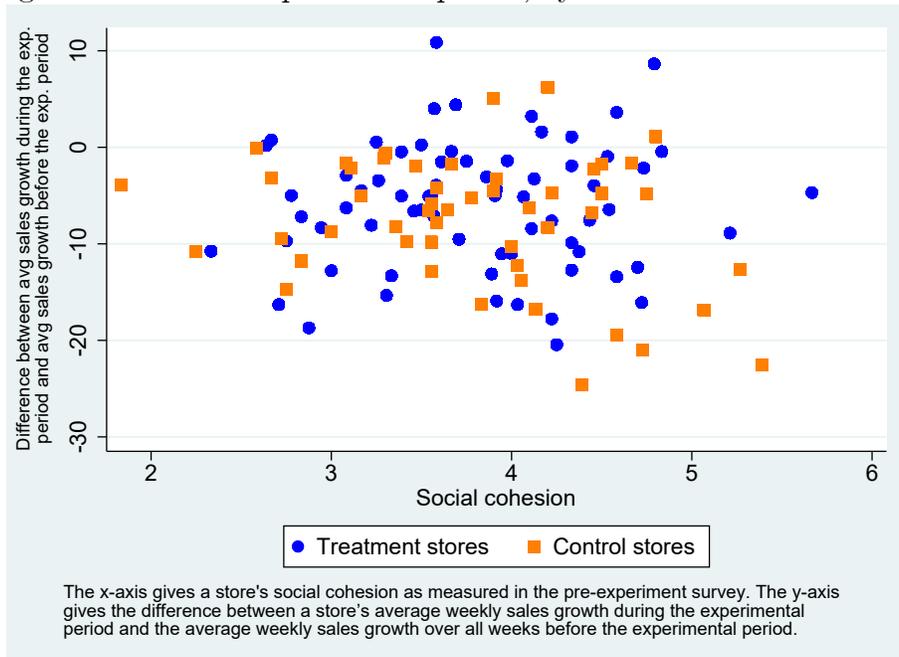
Figure 8: Estimated treatment effect as a function of social cohesion

The solid line depicts the estimated treatment effect, as estimated by an OLS fixed effects regression where both the treatment dummy and the experimental period dummy are interacted with a quadratic polynomial of stores' pre-experiment social cohesion.
The dashed lines represent the 95% confidence interval.
The gray area depicts the density of social cohesion across stores.
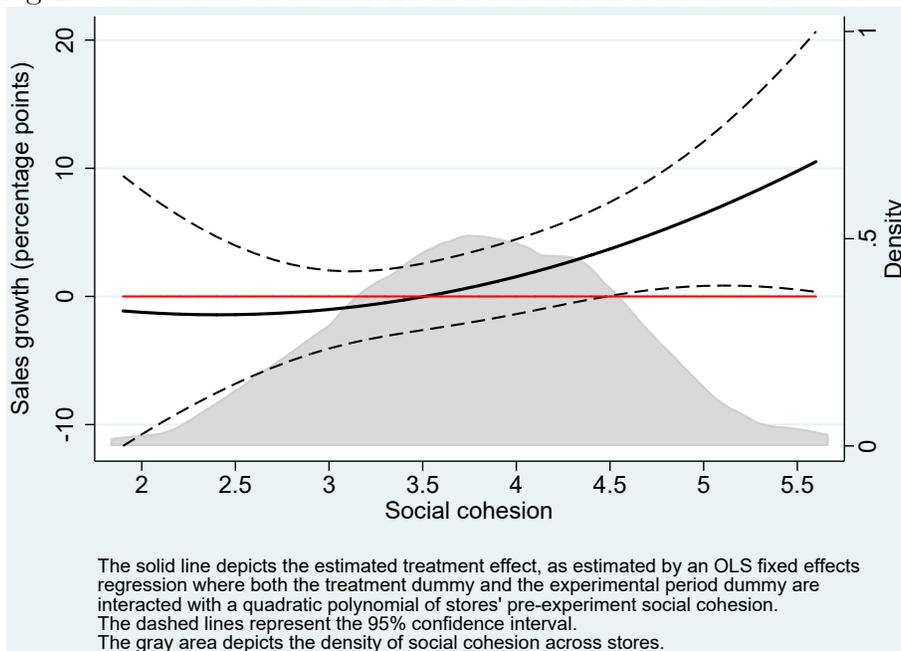
Table 1: Survey questions on social cohesion, adapted from Carless and De Paola (2000)

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 'Completely disagree'.

Table 2: Administrative data descriptives at the store level

	All stores		Control group		Treatment group	
	Mean	Std	Mean	Std	Mean	Std
Sales performance (store averages)						
Avg weekly sales growth before experimental peiod						
(week 1 - week 41 in 2013)	0.26	9.02	-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
Personnel characteristics (at the store level)						
Number of employees	9.87	2.71	9.78	2.45	9.93	2.90
Fraction of male employees	0.15	0.13	0.15	0.14	0.14	0.12
Avg tenure of employees (years)	5.99	2.61	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 of store manager (male = 1, female = 0)	0.58	0.49	0.56	0.49	0.59	0.49
Tenure of store manager (years)	13.98	8.71	14.11	8.01	13.89	9.23
Age of 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 at the store level

	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
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
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 percent level,
respectively, between control and treatment stores (t-test).
All survey measures are on a 7-point Likert scale.

Table 4: Treatment effects on sales growth

	Dependent variable: sales growth (percentage points)			
	(1)	(2)	(3)	(4)
Treatment	1.495	1.628	1.411	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
			(1.929)	(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)	(1.223)
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
				(0.600)
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 continuous interaction variables are mean-centered.

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

Table 5: 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)	(0.067)	(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 continuous interaction variables are mean-centered.
***, **, * denote statistically significant effects at the 1%, 5% and 10% level, respectively.

Table 6: Robustness checks

	Dependent variable: sales growth (percentage points)					
	Placebo		No store-fixed effects		Social cohesion excl. managers	Controlling for dispersion in social cohesion
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.717	-0.555	1.536	1.669	1.676	1.590
	(1.930)	(1.942)	(1.223)	(1.212)	(1.243)	(1.225)
Treatment X Social cohesion		1.372		2.851*	2.817*	2.650
		(2.229)		(1.685)	(1.697)	(1.941)
Exp. period X Social cohesion		-2.228		-2.350*	-1.877	-2.666*
		(1.763)		(1.276)	(1.282)	(1.472)
Treatment group			1.915	2.118		
			(1.618)	(1.640)		
Social cohesion				-0.343		
				(1.564)		
Treatment group X Social cohesion				-0.368		
				(1.951)		
Exp. period X st.dev Social cohesion						0.976
						(2.174)
Treatment X st.dev Social cohesion						1.187
						(3.527)
Store-fixed effects	yes	yes	no	no	yes	yes
Week-fixed effects	yes	yes	yes	yes	yes	yes
Store-week observations	4990	4908	5722	5628	5534	5346
Number of stores	122	120	122	120	118	114
R^2	0.646	0.645	0.585	0.585	0.635	0.633

Standard errors clustered at the store level in parentheses.

Note that Treatment is defined as Treatment group X Experimental period.

st. dev Social cohesion is the standard deviation of social cohesion among employees wthin the store.

All continuous interaction variables are mean-centered.

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

A Appendix

Figure A1: Average weekly sales growth by treatment status

Average sales growth is the percentage change in weekly sales at the store level
compared to the same week in the previous year, averaged over all stores in
either the treatment group of the control group.
The experiment took place in the period between the vertical red lines (weeks 42 to 47).
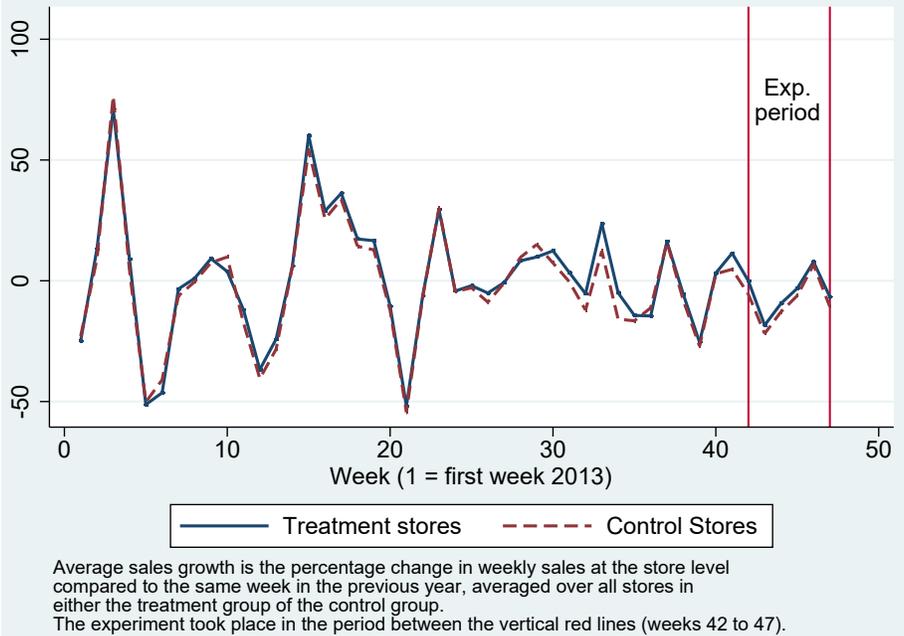
Figure A2: Average weekly sales growth by treatment status and social cohesion quintile
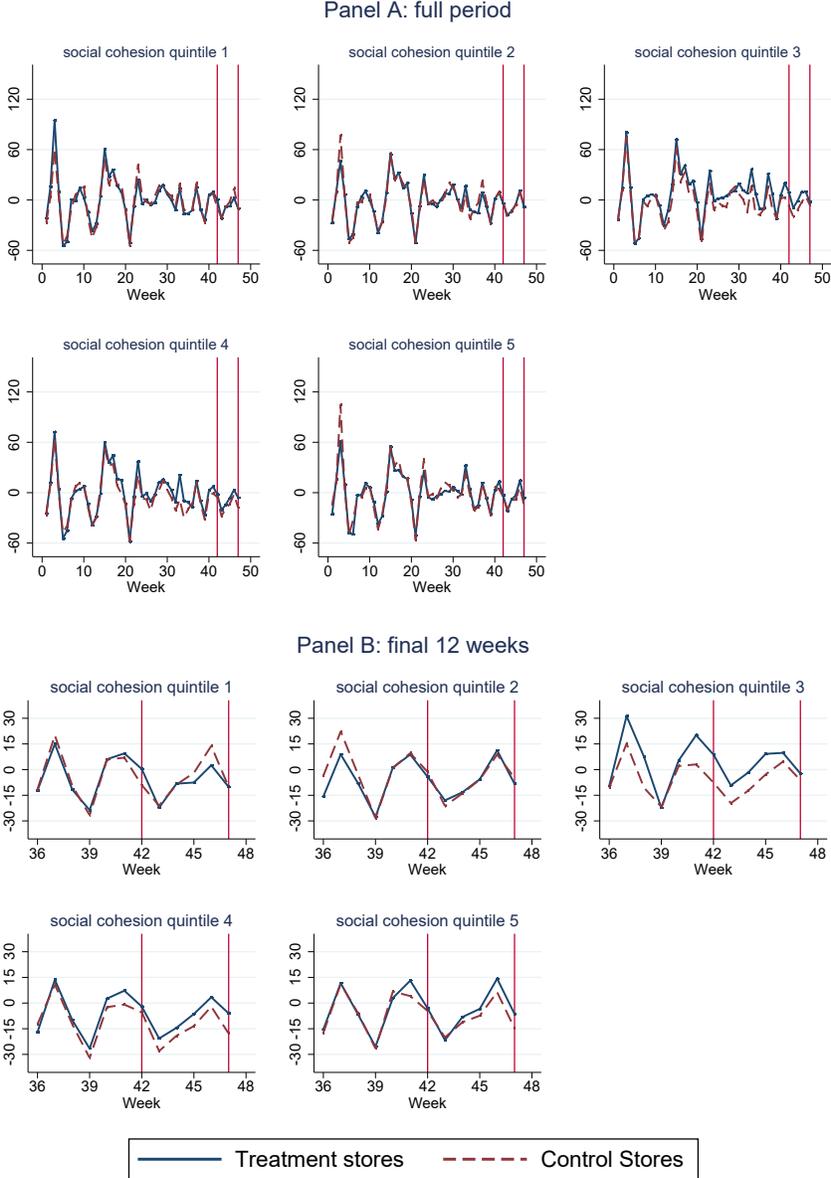
Stores are grouped into social cohesion quintiles based on the pre-experiment survey. The graphs show average sales growth by week and treatment status. The experiment took place in the period between the vertical red lines (weeks 42 to 47).

Figure A3: Actual and placebo average treatment effect by social cohesion quintile
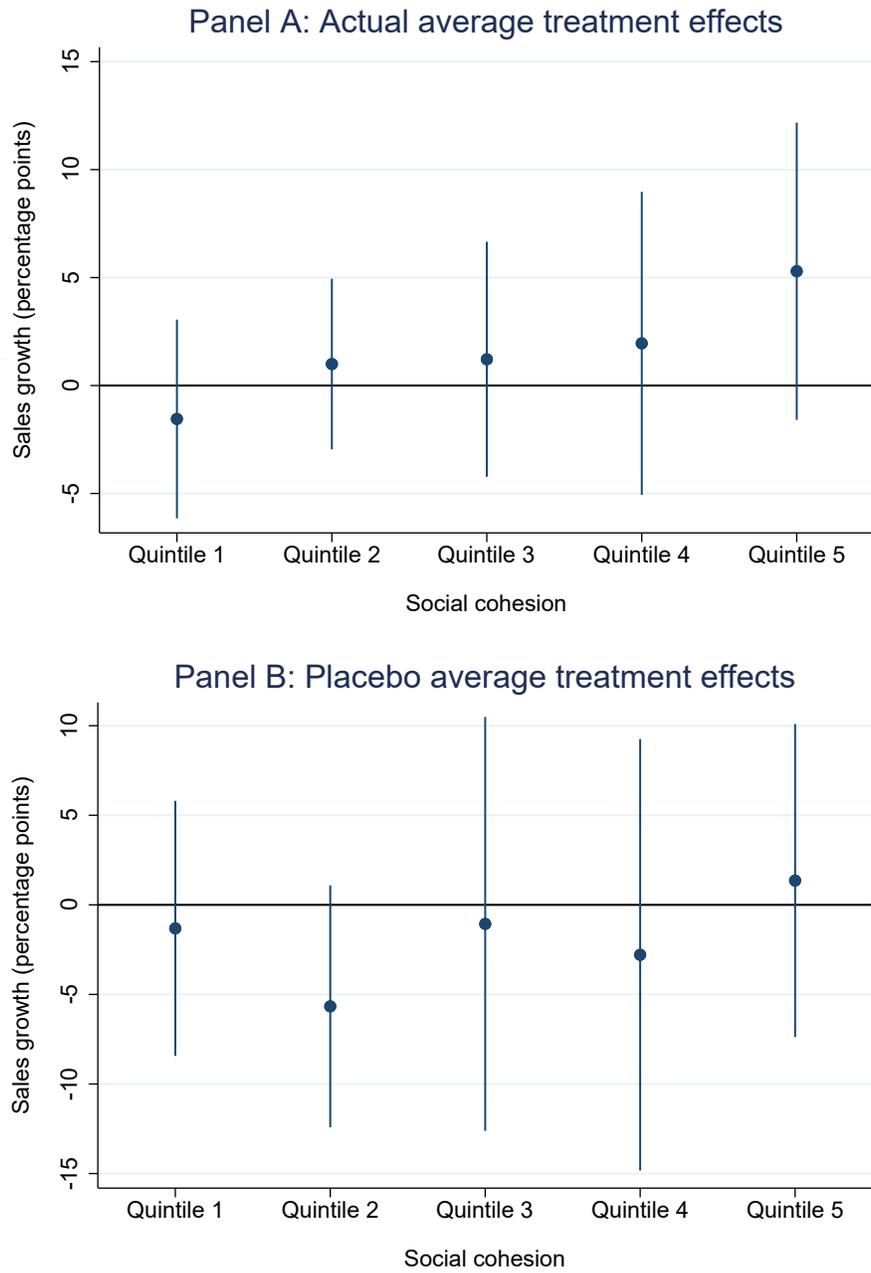
Panel A depicts the average treatment effect, estimated using equation (1), by social cohesion quintile. Panel B depicts the placebo average treatment effect by social cohesion quintile, estimated as if the experiment took place in the six weeks before the actual experiment. The lines represent the 95% confidence interval.

Table A1: Correlations between store-level characteristics

	Avg sales growth before exp.	Social 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												
Social cohesion	-0.04	1											
Task cohesion	-0.04	0.31	1										
Directive leadership	-0.11	-0.14	-0.07	1									
Supportive leadership	0.14	0.34	0.30	-0.18	1								
Helping	-0.02	0.39	0.64	-0.15	0.47	1							
Peer pressure	-0.05	-0.29	-0.15	0.04	0.00	-0.04	1						
Number of employees	0.05	0.12	-0.03	-0.05	-0.12	-0.06	-0.06	1					
Average tenure employees	-0.11	0.06	-0.17	0.06	0.17	-0.20	-0.09	0.01	1				
Average age employees	-0.04	-0.04	-0.11	0.01	0.11	-0.19	-0.04	-0.06	0.74	1			
% male employees	-0.07	0.00	0.07	-0.06	0.01	0.05	-0.09	0.11	-0.14	-0.04	1		
% 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	
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
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 the correlations with peer pressure, which are based on 117 stores.

Table A2: 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
Observations	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		
Observations	302 - 326	302 - 326		

Observations are the number of employees (columns 1 to 3) and stores (column 4) with responses.

These numbers vary per variable due to item non-response, in particular for peer pressure.

The correlations between pre- and post-experiment measures are calculated using only the stores in the control group, as the post-experiment measures in the treatment group may have been affected by the team incentive.

Table A3: Questionnaire data: respondents versus 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
Pre-exp. questionnaire												
Manager												
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
Number	94		26		39		11		55		15	
Employees												
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	7.71	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
Number	380		704		149		290		231		414	
Post-exp. questionnaire												
Managers												
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	0.47	0.51	0.62	0.49	0.44	0.51	0.63	0.49	0.50	0.51
Number	80		40		34		16		46		24	
Employees												
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
Number	246		838		88		351		158		487	

The personnel variables are extracted from the company's database as of October 2013.
Two stores did not have a store manager, implying that the total number of managers in our sample is 120.
***, **, * denote statistically significant differences at the 1%, 5% and 10% percent level, respectively, between respondents and non-respondents within a questionnaire (t-test, non-parametric (Wilcoxon rank-sum) test yields same result).
None of the differences between respondents and non-respondents are statistically significantly different between the control group and the treatment group at the 10% level.

Table A4: Descriptive statistics separately for stores with high and low social cohesion in the pre-experiment survey

| | Below-median social cohesion | | | | Above-median social cohesion | | | |
| | Control | | Treatment | | Control | | Treatment | |
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Performance(store averages)								
Avg weekly sales growth before exp. period (week 1 - week 41 in 2013)	-0.55	5.73	2.26	11.39	-1.20	11.08	0.23	6.38
Avg weekly sales growth during exp. period (week 42 - week 47 in 2013)	-6.66	3.98	-2.81	13.30	-10.16	14.55	-6.54	9.43
Within-store st.dev. of sales growth before exp. period	29.41	5.94	30.65	9.09	31.91	9.83	29.80	4.05
Store characteristics								
Number of employees	9.36	2.14	10.14	3.12	10.20	2.71	9.94	2.61
Fraction male employees	0.14	0.13	0.17	0.13	0.17	0.16	0.12	0.09
Avg tenure of employees (years)	6.57	2.88	5.54	2.52	6.02	2.69	5.99	2.53
Avg age of employees	28.40	5.18	27.36	3.96	26.01	3.17	27.13	5.08
Fraction on-call employees	0.41	0.14	0.42	0.14	0.42	0.12	0.39	0.14
Gender store manager (male = 1)	0.62	0.48	0.62	0.49	0.50	0.50	0.57	0.49
Tenure store manager (years)	15.04	8.22	13.18	9.92	13.09	7.83	14.71	8.68
Age store manager	42.98	10.92	38.44*	9.07	38.37	7.19	40.21	9.32
Pre-experiment survey								
Response rate employees	0.33	0.20	0.36	0.19	0.35	0.15	0.35	0.17
Response manager	0.76	0.44	0.80	0.41	0.85	0.35	0.78	0.41
Social cohesion	3.18	0.50	3.27	0.39	4.40	0.44	4.36	0.40
Task cohesion	4.62	0.66	4.69	0.76	4.77	0.58	5.03	0.63
Directive leadership	4.53	0.60	4.58	0.70	4.33	0.42	4.44	0.49
Supportive leadership	5.13	0.86	4.97	0.95	5.44	0.86	5.67	0.65
Helping	5.72	0.54	5.64	0.69	5.96	0.54	6.10	0.50
Peer pressure	4.87	0.69	4.98	0.72	4.69	0.70	4.71	0.80
Job satisfaction	5.67	0.69	5.48	0.75	5.63	0.71	6.03**	0.45
Number of stores	25		35		25		35	

The personnel variables are extracted from the company's database as of September 2013. All survey measures are on a 7-point Likert scale. Stores are grouped into below-median and above-median social cohesion based on the pre-experiment survey. ***,**,* denote statistically significant differences at the 1%, 5% and 10% level, respectively, between control and treatment stores (t-test).

Table A5: 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 of employees	0.084	(0.055)
St.dev. tenure of employees	-0.046	(0.048)
Average age of employees	-0.022	(0.029)
St.dev age of employees	-0.019	(0.034)
Fraction of male employees	0.448	(0.869)
St.dev. gender of employees	-0.515	(0.791)
Fraction of on-call employees	-0.376	(0.677)
Tenure of manager	0.022*	(0.011)
Age of manager	-0.008	(0.010)
Gender of manager (male=1)	-0.081	(0.183)
Stores	116	
R^2	0.288	

Standard errors in parentheses.

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

Table A6: Additional robustness checks

	Dependent variable: sales growth (percentage points)						
	Before-After		Clustering at group level		Winsorizing		Controlling for Job satisfaction
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	1.520	1.650	1.495	1.628	1.412	1.552	1.838
	(1.216)	(1.212)	(1.299)	(1.279)	(1.188)	(1.181)	(1.240)
Treatment X Social cohesion		2.782		2.891*		2.627	3.274*
		(1.683)		(1.730)		(1.602)	(1.743)
Exp. period X Social cohesion		-2.252*		-2.365*		-2.196*	-2.401*
		(1.272)		(1.284)		(1.163)	(1.271)
Treatment X Job satisfaction							0.797
							(1.761)
Exp. period X Job satisfaction							-1.611
							(1.223)
Store-fixed effects	yes	yes	yes	yes	yes	yes	yes
Week-fixed effects	yes	yes	yes	yes	yes	yes	yes
Store-week observations	244	240	5722	5628	5722	5628	5586
Number of stores	122	120	122	120	122	120	119
R^2	0.513	0.523	0.638	0.638	0.002	0.002	0.640

Social cohesion and Job satisfaction are mean-centered. Standard errors are clustered at the store level in columns (1), (2), (5), and (6). In columns (3) and (4), standard errors are clustered at the competition group level for treatment stores, yielding 68 clusters.

In columns (1) and (2), the data at store level is collapsed to the average sales growth before the experimental period and the average sales growth during the experimental period, respectively.

In columns (5) and (6), the top and bottom 1 percent of observations of the dependent variable after controlling for week-fixed effects are winsorized.

All continuous interaction variables are mean-centered.

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