Incentives and Social Relations in the Workplace

ISBN: 978 90 3610 238 4

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 502 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Incentives and Social Relations in the Workplace

Prikkels en Sociale Relaties op de Werkplek

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus
Prof.dr. H.G. Schmidt
en volgens het besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
vrijdag 1 juli 2011 om 13:30 uur
door

Joeri Sol
geboren te Leidschendam

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor:	Prof.dr. A.J. Dur
Overige leden:	Prof.dr. D. Sliwka
	Prof.dr. B. Visser
	Prof.dr. H.D. Webbink

Preface

I enjoyed the four years that led to the thesis that you see before you. My thesis has benefitted a great deal from discussions with people from both inside and outside of academics. Here I take the opportunity to thank some people individually.

First and foremost, I want to thank Robert Dur. Robert, your enthusiasm in doing research is contagious, and your ingenuity is inspiring. I always left your office with renewed energy and motivation, while there were times that I was close to despair upon entering your office. Although I should probably discount some of the positive reinforcement, your positive attitude and encouragement made me more confident about my ability as a researcher. Besides being an academic role-model, you are also one of the most fun people to be around during dinners, after workshop drinks, and in the club.

Second, I like to thank my other co-authors in this thesis: Josse Delfgaauw, Willem Verbeke, and Okemena Onemu. Josse, you showed me how to structure my work better, and I enjoyed being the teaching assistant for your course. Willem, our joint company visits have been a success and working together was instructive. Oke, your dedication and eagerness to learn will leave you with a PhD soon enough.

Furthermore, I would like to thank Dirk Sliwka, Bauke Visser, and Dinand Webbink for their willingness to take part in my doctoral committee, as well as the members of the large committee, Mirjam van Praag and Otto Swank. In particular, I like to thank Dirk Sliwka for his comments. In addition, I want to express my gratitude to the management of the anonymous Dutch retail chain that is frequently mentioned in this thesis. A special thank you to Jerry and Martin for their patience and enthusiasm for working together with researchers.

On the 9^{th} floor I found plenty of inspiration for my topic: Social relations in

the workplace. I am thankful to have shared an office with two fellow students who started the same year. Bart and Paul, you have been helpful sparring partners in discussions about research and teaching. In addition, you and Alexei provided healthy distractions with chess, other board games, and office soccer. Arjan and Margaretha, sharing the same supervisor ensured that I always found an interested ear or someone to laugh with, even in Granada, Venice, or Barcelona. I enjoyed the chats over lunches or breaks with Riemer, Sven, Umut among many others from TI. I also enjoyed all social and professional interaction with the staff from the 8^{th} floor, in particular I want to thank Benoit Crutzen and Vladimir Karamychev for useful suggestions. I have found another nice group of colleagues at my new position with the University of Amsterdam; where I managed to finish my final chapter with some pointers from Jeroen van de Ven.

I moved to Rotterdam when starting the PhD. Bram, next to being an excellent roommate, master of the Dutch cuisine, talented basketball player, and fair loser with PES, I value your reliability and passion for your environment the most. I want to thank Elisha for good memories and for enduring countless practice presentations. I also want to thank the 'jongahs van de SV', as well as the Baros people, for making Rotterdam feel like home.

I want to thank friends and family that have always supported my studies. Ramzy, discussions about my work and your business, combined with your sense of humour, give a good reason to visit Zoetje. Most importantly, Zoetermeer is the place where I find a warm and loving home. André, Petronella, and Mario, I cannot thank you enough for your unconditional love and support. Finally, during the last year of my PhD I met Aslı. Aslıcım, thanks you for helping me through the last bits of this thesis, and for being able to make me forget about work altogether.

Contents

Chapter 1

Introduction

"Interdependence is and ought to be as much the ideal of man as self-sufficiency. Man is a social being." Mohandas K. Gandhi (1929)

A central theme in personnel economics is the misalignment of the goals of employees and those of the organization that employs them. In case the required actions are impossible or costly to verify, this misalignment gives rise to the well-known problem of moral hazard (Holmstrom 1979). Organizations have to ensure that it becomes in the employee's best interest to work towards the goals of the organization. A widely studied means of achieving this goal alignment is making an employee's income contingent on objectives set by the employer; e.g., a piece rate, an annual bonus depending on performance, or a promotion to reward outstanding behavior over a longer period of time. There is compelling evidence that these financial incentives increase employee productivity (see Prendergast 1999 and Lazear and Oyer 2009 for an overview).

While financial rewards can be a powerful tool to promote alignment, it is by no means always the only, or most efficient, option to motivate employees. Some organizations can provide employees with other highly valued job aspects, such as intrinsically rewarding work (Delfgaauw 2007), social inclusion, or an inspiring boss (Dur et al. 2010). Moreover, neither financial incentives nor non-monetary rewards should be studied in a vacuum, as important interaction effects may occur. In this thesis, I study how different workplace practices influence, and are influenced

by, interpersonal relationships between colleagues. By enriching economic theory in this direction I aim to increase the understanding of organizational design and human behavior in organizations. Further, this thesis may offer organizations some lessons in their quest to optimize human resource practices.

This thesis presents two theoretical chapters and two chapters that describe the results of field experiments. I provide theoretical investigations into the interaction between financial incentives and co-worker relations using principal-multi-agent models. Principal-multi-agent models capture the strategic interaction between an organization and its employees (such as the moral hazard problem described above), and allow for a formalization of interpersonal relationships between colleagues. Further, I present the results of two field experiments on the effects of team incentives. Empirical evidence for interaction effects between financial incentives and co-worker relations is difficult to obtain, because it is hard to establish causality with naturally occurring data. For example, when we observe a positive correlation between the quality of co-worker relations and the use of team incentives, it is unclear whether good co-worker relations are a determinant for the use team incentives or a result of it. Field experiments circumvent this problem of reverse causality by randomization.

This introduction proceeds as follows: The importance of co-worker relations is motivated in Section 1.1. I will provide a more general discussion of the merits and caveats of field experiments in Section 1.2. The introduction ends with an overview of the different chapters in Section 1.3.

1.1 Co-worker Relations

Most organizations actively promote social interaction among their personnel; examples range from the facilitation of coffee corners to offering joint holiday trips (e.g., Rohlen 1975, Cohen and Prusak 2001). There are two main reasons for firms to invest in social interaction between colleagues: First, productivity may be higher when employees have good relations on the workfloor. Second, there may be a compensating wage differential for good co-worker relations, i.e., employees may be willing to accept lower wages in workplaces with a more pleasant atmosphere.

The belief that co-worker relations can be of value to an organization is shared among managers. A survey among managers in the public sector in the US revealed that close to 85% of managers state that their organization actively promotes friendships between colleagues (Berman et al. 2002). These managers observe higher productivity in workplaces where friendships were stimulated. However, evidence for a direct relationship between group cohesiveness and productivity is inconclusive (see the studies discussed in Rotemberg 2006). Co-worker relations are found to be directly related to other important measures; co-worker relations are positively correlated with job satisfaction, organizational commitment, while they are negatively related to employee stress, absenteeism, and turnover (see Price and Mueller 1981, Riordan and Griffeth 1995, Nielsen et al. 2000, and Morrison 2004, among other studies mentioned in Chapter 2).

Facilitation of social interaction among colleagues makes firms more attractive as an employer for many (potential) workers. The anticipated interaction with colleagues is one of the main drivers for job search among Dutch unemployed (Van Echtelt and Hoff 2008). Besides this, retired workers state that the social interaction with colleagues is one of the most missed job aspects (Shacklock 2005). Finally, detailed time-use data from both France and the US illustrate that employees are more satisfied and in a better mood, when their job involves more frequent interaction with colleagues (Krueger and Schkade 2009). The theory of equalizing differences (Rosen 1986) predicts that workplaces that have better co-worker relations are associated with a lower wage. Borzaga and Depedri (2005) provide evidence for this negative relation between co-worker satisfaction and wages in the Italian non-profit sector. In line with this, Hamilton et al. (2003) show that some employees in a Californian garment factory voluntarily agreed to join team production, even though this resulted in an income loss, suggesting that these workers received some non-monetary rewards from teamwork.

This thesis studies the interaction between co-worker relations and the financial incentives that are offered. Besides a direct investment in the social interaction between co-worker relations, firms may have some influence over co-worker relations with the financial incentives they offer their employees. Employees may treat their

colleagues differently depending on how their colleagues' effort affects wages. The existing body of literature identifies a rich number of channels through which a worker's compensation can shape his actions towards colleagues: Team incentives may stimulate helping behavior (Fitzroy and Kraft 1986 and Itoh 1991a), or lead to increased peer pressure (Kandel and Lazear 1992 and Barron and Gjerde 1997). Promotion tournaments may induce colleagues to sabotage each other (Lazear 1989). This thesis contributes to the literature that studies how incentives influence the interaction between colleagues, with both a theoretical model in Chapter 2 and a field experiment in Chapter 4.

Notwithstanding the positive aspects of social interaction between co-workers mentioned above, facilitating interaction is not without risk. Managers perceive office gossip, distraction from work, and disturbance of merit-based decision making, as threats that come with workplace friendships (Berman et al. 2002). Chapter 3 of this thesis studies the latter concern that interpersonal relations between colleagues may influence the effectiveness of the financial incentives that are being offered.

Social relations in the workplace influence how employees react to the incentives in place, and thereby shape the opportunities to introduce effective financial incentives. For example, employees with good co-worker relations may take the effect of their effort on colleagues more into account (Rotemberg 1994). Bandiera et al. (2005) found that productivity went up when a firm moved from relative performance incentives to individual piece rates, especially for those workers who worked alongside their friends. In line with this, in a one-time employment setting, Cohn (2010) found that a negative externality in wages led to a significant drop in productivity when employees were given the opportunity to interact with co-workers. In this study, social interaction did not increase productivity under team incentives. Further, in a laboratory experiment, Towry (2003) shows that a team incentive was strengthened with team identity, whereas productivity was undermined when compensation depended on the report of a colleague.

1.2 Field Experiments

According to empiricists, knowledge arises from experience and observation. This notion is central to the practice of science, where theories have to withstand empirical scrutiny. However, simple observation is often not sufficient to falsify a hypothesis, as one needs the counterfactual of an observation to find conclusive evidence. For example, to conclude that a policy had a positive impact on worker productivity, it is not enough to observe an increase in productivity. There may have been factors unrelated to this policy that have led to the increase. Ideally, one would like to observe the same workers in the scenario where they were not exposed to the policy, i.e., the counterfactual. It is the empirical economist's task to create such counterfactuals.* Experiments create counterfactuals by random assignment to a policy; workers that are assigned to the control group serve as the counterfactual to those that are assigned to the policy. With naturally occurring data, the empirical economist has to use some identification method to create the counterfactual, e.g., with instrumental variables, or matching. Field experiments are the latest addition to the empirical economist's toolbox. In this section, I discuss some vices and virtues of field experiments: It is not meant to be exhaustive, for a more detailed overview of field experiments in labour economics see List and Rasul (2010).

According to John List, an advocate of field experiments in economics, field experiments build a bridge between laboratory experiments and naturally occurring data (List 2006). The credibility of the results in studies using naturally occurring data, i.e., the internal validity, depends on the quality of the identification method that has been used. Laboratory experiments directly allow for causal inference, as randomization of the treatment gives the counterfactuals in the control group. However, laboratory experiments may lack realism in the stakes involved, the representativeness of the participants, or the presence of an experimenter, among other dimensions (Harrison and List 2004, Levitt and List 2007). Therefore, its external validity is often questioned. A natural field experiment, defined as an experiment

*Other goals of empirical research can be to make inference about a larger population from observations in a smaller sample (empirical exploration), or to estimate or forecast variables of interest.

where participants act in their natural environment without being aware of the experiment (Reiley and List 2008), combines the virtues of randomization and realism.

Obviously field experiments are no magic potion. When moving to the field, experimenters often have to give up some control in exchange for a more realistic setting. There also remain concerns with the external validity, as results from field experiments may be specific to the local conditions (examples abound in Rodrik 2009). Even if the results from small-scale field experiments would not be specific to local conditions, the results do not automatically generalize to a larger population due to possible general equilibrium effects. For example, suppose a field experiment has shown that more productive workers sort into a firm that pays for performance, such sorting is unlikely to occur when all firms start paying for performance. Most discussions of the merits and limitations of field experiments conclude that they complement the existing empirical methods.

Finally, there are some concerns that researchers should keep in mind irrespective of the empirical method that is being applied. A first concern is the sample selection bias, meaning that the studied sample differs from the population at large in ways that may bias the results. The results in Chapter 4 and 5 may also suffer from this bias. Namely, the field experiment took place in only one of the 15 companies that we approached. Therefore, it is possible that this company had some distinct features, like a management that was willing to experiment. This may have had an influence on the results. However, the sample selection bias is not specific to the experimental method alone; naturally occurring data may also suffer from this bias. Banerjee and Duflo (2009) provide the following example: "Large programs are politically more sensitive to evaluate than pilot programs, since they are usually well publicized, and so countries may be strategic with respect to the choice of programs to evaluate." (p. 19).

A second concern is the publication bias, which arises when the likelihood of research findings being published depends on having results that are significantly different from zero.[†] The non-reporting of zero results is also referred to as the "file

[†]There may be good reasons for journals to favor studies with significant results, as their research method or data may have been of superior quality. For example, Chapter 4 will probably have difficulties ending up in a journal, as there are multiple viable explanations for the zero

drawer problem", where studies without significant results proverbially disappear into the file drawer (Rosenthal 1979). Rosenthal argues that this problem can be overcome by replication, as a large number of studies showing significant results would require an inconceivable large file drawer with zero results to counteract the existing results. However, the combination of theoretical presumptions among researchers and the discretion they have over the control variables they include, can still lead to a harmful publication bias. Card and Krueger (1995) argue that the literature on the effect of minimum wages on employment may suffer from such bias. In a meta-analysis they show that the estimated effect typically is twice the standard error, irrespective of the size of the effect, and that the t-values do not increase with the number of observations. Finally, the publication bias may not be limited to the favoring of significant results. Replications of earlier studies have difficulties making their way into leading journals even when their findings are significant. This problem is arguably most severe for experiments. As Rodrik (2009) puts it: "Perhaps ironically, other types of studies that have weaker internal validity generate much greater incentive for replication. Here the name of the game is improved identification, ..." (p. 23).

1.3 Overview

The following two chapters provide a theoretical investigation of the interaction between a worker's financial incentives and the interpersonal relations between co-workers. Chapter 2 explores the possibility to stimulate co-worker relations by fine-tuning a worker's financial incentives. In the principal-multi-agent setting under study, agents not only choose productive effort, but also engage in social interaction with their colleagues. Social interaction is modeled as an exchange of 'attention', where giving attention is costly. The receipt of attention gives a consumption benefit and affects a worker's social preferences: That is, agents are conditionally altruistic, i.e., altruism towards a colleague increases in the receipt of his attention. The private costs and external benefits of attention provision give rise to an internalization

result. However, the publication bias implicitly refers to favoring studies with significant results over studies with zero results of comparable research quality.

problem, resulting in suboptimal attention provision.

We show that the principal can stimulate social interaction between colleagues by offering agents either team incentives or relative performance incentives. This externality in a worker's compensation gives agents the desire to influence the effort choice of their colleagues. The strategic motivation to provide additional attention to a colleague is the following: As the co-worker's altruism towards the giving agent is increasing in the receipt of attention, attention provision will trigger a reciprocal reaction. That is, the altruistic colleague partly internalizes the effect of his effort on the giving agent, adjusting his effort in the desired direction in response to the receipt of attention. Thereby, the principal intentionally introduces an externality of effort to mitigate the externality problem of attention provision. The principal restores incentives for effort either by muting individual incentives for effort in combination with team incentive, or amplifying individual incentives when combined with relative performance incentives, such that the first-best is obtained.

Chapter 3, instead, studies how existing co-worker relations influence the optimal use of incentives. I look into a setting where co-workers have better information about each other's behavior than their manager. In the model this boils down to colleagues receiving signals about each other's effort choice, while the principal only observes the team output. The principal attempts to improve upon offering a team bonus by asking agents to send evaluation reports about their colleague. I start from ideal circumstances for these peer evaluations by assuming that agents have costs of lying about the signals they receive. Truthful peer evaluation gives an agent incentives to exert effort, because he desires to increase the likelihood that his colleague receives a positive signal and evaluates performance accordingly. In this case, a bonus for the receipt of a positive peer evaluation performs just as well as individual incentives for effort.

Next, I allow for co-worker relations, either good or bad, which may bias the evaluation decisions. Agents trade-off the cost of lying and the internalized utility of a bonus given to a (dis)liked colleague. Small costs of lying, strong social preferences, or a large reward for a positive evaluation can lead friends to give each other positive evaluations irrespective of the received signal, as well as cause foes

always to begrudge each other the peer evaluation bonus. The principal can ensure that peer evaluation remains truthful by adjusting the peer evaluation bonus downward. Nevertheless, the optimal contract always includes a peer evaluation bonus, which is complemented with a team bonus in case peer evaluation becomes severely restricted.

Chapter 4 and 5 describe the results of two field experiments. Both experiments took place in a Dutch retail chain of 128 stores. The stores operate under a single brand, and sell clothing and shoes. The management wanted to introduce incentives for store employees, who, unlike store managers, only receive a fixed wage. In both experiments we introduced a temporary team incentive for all employees within a store. Neither store managers nor store employees were aware that they took part in an experiment, which classifies the experiments as natural field experiments (Harrison and List 2004).

Chapter 4 studies the determinants and the effects of anti-shirking behavior; i.e., actions that employees undertake when they observe a colleague not working as hard as he or she should. During the experiment all employees in treated stores could earn a team reward by raising the number of items sold per purchase above a given target. We hypothesized that anti-shirking behavior may influence the effectiveness of a team incentive and may be affected by the presence of a team incentive. Therefore, we held a questionnaire to gauge a store's anti-shirking culture both prior to and after the experiment.

The team incentive did not lead to additional sales during the experiment. The short-run team incentive did cause a reduction in the willingness to undertake anti-shirking behavior. We provide suggestive evidence for the claim that this decrease is the result of dissatisfaction with increased peer pressure during the experiment. Namely, respondents in treated stores stated significantly more often that anti-shirking behavior led to resentment by the colleague at which the actions were aimed, and significantly less often that it led to an improvement of the colleague's behavior. Besides the unsatisfactory experiences with anti-shirking behavior, difference-in-difference estimations also show that co-worker relations suffered in treated stores. An alternative explanation, that employees might have showed a

general dissatisfaction in treated stores, is ruled out by showing that job satisfaction and employee-management relations are unaffected by the treatment.

The field experiment described in Chapter 5 introduced sales contests, i.e., tournaments, between stores. In one treatment the winning store and the runner-up could earn a financial reward, while in a second treatment there was only the honor of winning (i.e., no financial reward). Both types of tournaments led to significantly higher sales growth compared to the control group, and we find no difference between the treatments. This result is in line with recent literature that shows positive effects of relative performance feedback (Blanes i Vidal and Nossol 2009 and Azmat and Iriberri 2010) and non-monetary rewards of winning a contest (Kosfeld and Neckermann 2011).

In addition, Chapter 5 contributes to the literature on gender differences in competition, by exploring whether the treatment effects are heterogeneous in gender. Unlike recent findings (Gneezy et al. 2003, more references can be found in Chapter 5), we did not find that the response towards competition differed by gender, neither at the store manager level nor at the employee level. Interestingly, the interaction between the store manager and the gender composition of a team mattered for the effect of these competitions. Namely, the competition only had an effect in stores where the manager and a large fraction of the employees were of the same gender. Our findings are important for two reasons: First, we show that competition can be equally stimulating to women, which possibly depends on the way that the tournament is communicated. Second, the interaction is interesting for the competition over executive-level positions. In order to reach top positions, one should win several promotion tournaments, where performance depends not only on one's own performance, but also on the performance of the team one leads.

Finally, I conclude with Chapter 6, which provides a discussion of the results, some managerial implications, and suggestions for further research.

Chapter 2

Social Interaction, Co-Worker Altruism, and Incentives*

Joint with Robert Dur

2.1 Introduction

Social interaction with colleagues is a highly valued job aspect for many workers. Research in psychology, sociology, and management shows that receiving affective support from colleagues and having good interpersonal relationships at work are positively associated with job satisfaction, job involvement, and organizational commitment, and negatively with employee stress and absenteeism (see among others Price and Mueller, 1981; Riordan and Griffeth, 1995; Hodson, 1997; Ducharme and Martin, 2000; Nielsen et al. 2000; Morrison 2004; Wagner and Harter, 2006). Moreover, turnover intentions and actual turnover tend to be lower when workers experience social support from co-workers (Price and Mueller, 1981; Riordan and Griffeth, 1995; Nielsen et al., 2000; Morrison 2004; Mossholder et al., 2005). Social interaction with colleagues is also one of the most missed job aspects under retired workers in Australia (Shacklock, 2005) – and it is one of the main drivers of job search among Dutch unemployed (Van Echtelt and Hoff, 2008). Lastly, using time-

*A slightly adapted version of this chapter appeared in *Games and Economic Behavior* (2010) Vol. 69: 293 - 301.

use data for France and the US, Krueger and Schkade (2009) show that workers who are in jobs that entail more frequent interactions with co-workers are more satisfied with their jobs and in a better mood during work time.

These findings have a clear managerial implication: In their struggle to attract and retain workers, managers should strive to create and maintain high-quality co-worker relationships. This view is confirmed by managers. Berman et al. (2002) report the results of a survey among managers in the US and show that more than 85% of managers approve or strongly approve of workplace friendships. A similar percentage reports that their organization actively encourages workplace friendship. An obvious and widely used means of doing so is to facilitate social interaction among co-workers through e.g. creating coffee corners or a nice canteen, having Friday-afternoon drinks, or organizing after-work social events (Cohen and Prusak, 2001). However, as we shall see, when a company's workplace policies are limited to facilitating social interaction, typically too little social interaction takes place, implying lower than first-best profits.

This chapter studies an alternative, complementary way to promote co-worker relationships: fine-tuning workers' financial incentives. We develop a principal-multi-agent model in which workers do not only engage in productive activities, but also in social interaction with their colleagues. Workers' productive activities are, for convenience, assumed to be fully contractible. Social interaction, however, is not contractible at all. We model social interaction as an exchange of 'attention' between workers. Attention may include showing interest in a colleague's personal life, offering a drink after working hours, or any other kind gestures. While receipt of attention is always valued positively by workers, giving attention is assumed to be costly, at least above a certain level of attention. The reason is clear: Although giving some attention can evidently be pleasurable, it is also time-consuming, expensive, or perhaps even boresome at some point. In addition to these direct benefits and costs, we assume that social interaction creates altruistic feelings among colleagues. More specifically, we assume that receipt of attention leads to stronger feelings of altruism towards the giving agent. As we shall see, in equilibrium this gives rise to reciprocal behavior: When a worker has been treated kindly by a colleague, the

worker cares more about his colleague's well-being, and adapts his future actions accordingly.

We obtain two main results. First, when the firm provides only individual performance incentives, too little social interaction takes place, implying lower than first-best profits for the firm.[1] The reason is an externality problem. Each worker internalizes the benefits of giving attention to his co-workers in as far as he is altruistic towards his co-workers. Since people care more for themselves than they do for their colleagues, there is too little social interaction in equilibrium. This is costly to the firm: If the firm could induce workers to engage in more social interaction, workers' job satisfaction would be higher, allowing the firm to pay lower wages. Borzaga and Depedri (2005) have recently provided some evidence for such compensating wage differentials. They find that, in Italian non-profit organizations, satisfaction with colleagues is negatively associated with wages. Consistent with this, the field study by Hamilton et al. (2003) shows that quite a few workers of a Californian garment factory were willing to give up a substantial part of their salary so as to join team production, suggesting high non-pecuniary benefits from working in a team.

Second, the firm can promote social interaction among workers by including team incentives or relative incentives in the workers' contract. Consequently, the firm can achieve first-best profits by choosing the right mix of individual incentives and team or relative incentives. The intuition behind these results is as follows. Provision of team or relative incentives creates externalities among workers. Team incentives create positive effort externalities, implying underprovision of effort (free-riding); relative incentives create negative effort externalities, resulting in overprovision of effort from the perspective of the workers. These externality problems are less severe when workers are more altruistic towards each other. Hence, contracts with team or relative incentives strengthen workers' incentives to invest in co-worker altruism. A natural way to do so is to engage in social interaction with colleagues. In other words,

[1] A similar result can be found in Itoh's (1991b) study of social relations and incentive contracts, of which we became aware only after completing a first draft of this paper. In contrast to our model, workers in his model are not altruistic. Consequently, our results on optimal incentive contracts starkly differ from his.

by deliberately creating an additional externality problem among workers through provision of either team or relative incentives, firms induce workers to resolve the initial externality problem of too little attention provision. Incentives for productive activities are restored through fine-tuning the level of individual incentives. Optimal contracts thus induce workers to exert first-best effort and to give first-best attention. Consequently, the firm achieves first-best profits.

Our model's predictions concerning the effect of team and relative incentives on the quality of co-worker relationships and workers' effort are supported by existing empirical findings. Firstly, there is evidence based on survey data. Burks et al. (2009) study social preferences of bicycle messengers in Switzerland and San Francisco and find that, compared to couriers in firms that pay for individual performance, couriers under team-based or hourly pay act more cooperatively in a prisoner's dilemma game and expect their co-workers to be more cooperative as well.[2] Moreover, the paper provides some evidence that these differences in social preferences are endogenous to the employer's choice of compensation scheme. Likewise, Carpenter and Seki (2011) find among Japanese fishermen that those who pool their catch at the end of day are significantly more altruistic towards each other than those who organized themselves into groups that do not pool. Heywood et al. (2005) analyze data from the German Socioeconomic Panel and find evidence for the view that promotions and profit-sharing are alternative means of generating cooperation among workers. Their measure of cooperation is workers' response to the question "Do you get along with your colleagues?", which is close in spirit to the quality of interpersonal relationships that we focus on. Heywood and Wei (2006) examine data from the National Longitudinal Study of Youth (US) and find that co-worker satisfaction is significantly higher for workers who recently received a promotion. No such relation is found between co-worker satisfaction and individual performance pay, profit-sharing, or the wage level.

A potential problem with evidence based on survey data is reversed causality: It may well be that high-quality co-worker relations are a determinant rather than

[2]Unfortunately (for our purpose), the study pools the data for couriers under team-based and hourly pay. About 70% of the subjects in this pool receive team-based pay, the remaining 30% are on hourly pay.

the result of team or relative incentives. Field experiments circumvent this problem. Rotemberg (1994) discusses the famous Hawthorne experiments (Roethlisberger and Dickson, 1939) where both workers' financial incentives and their opportunities for socializing were varied. The results suggest that team incentives encouraged friendship among workers and that this friendship was instrumental in raising output. More recently, the field study by Bandiera et al. (2005) compares fruit pickers' productivity under individual incentives and relative incentives. Their results strongly suggest that, when workers are paid on the basis of relative performance, they partially internalize the negative externality their effort imposes on others, especially when working alongside their friends. They do not find evidence for pure altruism, however. One reason for this could be that workers in their sample are hired on a daily basis, with no guarantee of further employment, giving little incentives to build up relationships.

This chapter proceeds as follows. In the next section we give a brief overview of related literature and discuss how our chapter contributes to it. Section 2.3 presents the model. In Section 2.4 and 2.5 we examine the case of perfect contractibility and the case of non-contractible attention, respectively. Section 2.6 concludes.

2.2 Related literature

This chapter builds on Rotemberg's (1994) seminal analysis of human relations in the workplace. He argues that, when workers' actions are strategic complements and workers are paid as a function of joint output, they may rationally choose to become altruistic towards each other. Altruism serves as a commitment device to exert more effort, which – due to the strategic complementarity of workers' efforts – induces co-workers also to exert more effort. This is in the worker's narrow self-interest because of the free-rider problem inherent in team incentives.

We differ from his analysis in three important respects. First, while Rotemberg studies the effect of team incentives in isolation, we derive the properties of first-best contracts which are shown to consist of a mix of different types of incentives. Second, in contrast to Rotemberg, strategic complementarity between workers' productive

actions is not a necessary condition for co-worker altruism to arise in our model. The reason is that we allow for a consumption benefit from social interaction at work, which is absent in Rotemberg. Last, and most important, we do not allow people to directly choose their altruistic feelings towards others. Instead, we assume that people may attempt to make others feel more altruistic towards them by being kind, showing attention, paying respect, offering favors, and so on. Thus, while as in Rotemberg an individual's altruistic feelings are endogenous in our model, the individual does not choose his feelings, but his feelings can be affected by other people's actions.

The way we model social interaction between workers and how it affects co-workers' altruism is close to the formalization of social ties in van Dijk and van Winden (1997). In their model, as in ours, social ties are represented by inter-dependent utility functions, where the weight assigned to the utility of the other agent depends on the history of interaction.[3] The positive relationship between social interaction and the formation of social ties is supported by a large number of empirical studies in several branches of the social sciences. For example, Homans (1950) concludes from observations of workers at the Western Electric Company that "favourable sentiments increase as interaction increases" (p. 112). Additional support for this hypothesis can be found in Baumeister and Leary (1995), van Dijk et al. (2002), and Hays (1988). We differ from van Dijk and van Winden (1997) in the application, as they analyze the influence of social ties on the contribution to a public good. Further, we do not make the assumption that individuals are myopic with respect to the feelings of a colleague; instead workers may invest in social relationships for strategic reasons, e.g., to alleviate externality problems.

In another related approach, Cox et al. (2007) have developed a model where the marginal rate of substitution between an agent's own income and the income of another is influenced by actions of this other agent. In particular, an agent becomes more willing to pay for the income of the other agent, i.e., becomes more altruis-

[3] In a related approach by Bolle and Kritikos (2006), the altruism parameter is not defined as the weight assigned to the utility of the other agent, but as the marginal utility of a transfer to another agent. However, like in van Dijk and van Winden (1997), this altruism parameter depends on the past interaction with this agent.

tic, if the other agent has been more generous to the former. Recently, Cox et al. (2008) formulated a similar theory in a more general (nonparametric) framework of preferences over one's own and other people's payoffs. Both papers discuss results of existing laboratory experiments to validate the model. The results of these experiments indicate that people do become more altruistic in response to kind behavior.

The results of our analysis are in stark contrast to those of Lazear (1989) on sabotage in tournaments and of Kandel and Lazear (1992) on peer pressure in teams. These papers predict *worse* rather than better co-worker relations under relative or team incentives compared to individual incentives (see also Barron and Gjerde, 1997). While we obviously do not deny that sabotage and peer pressure are relevant in many settings (see e.g., Garicano and Palacios-Huerta (2005) on 'dirty play' in professional soccer), the empirical evidence discussed in the previous section strongly suggests that opposite forces such as those studied in this chapter can sometimes dominate.

The economics literature provides two alternative ways through which team-based pay may improve upon the work climate: by increasing workers' willingness to help each other and by reducing pay inequity at the workplace. Studies stressing workers' helping behavior include FitzRoy and Kraft (1986), Drago and Turnbull (1988), Itoh (1991a), Rob and Zemsky (2002), and Corneo and Rob (2003). A crucial difference between these studies and ours is that helping or cooperating is assumed productive in these studies, implying that there is a team-element in production, which is not necessarily the case in our model. This chapter can thus explain the prevalence of team-based pay and their positive effects on the work climate, even when there is little or no complementarity between workers' productive efforts. The same holds for studies which consider inequity-averse workers (see Bartling, 2007; Demougin and Fluet, 2006; Englmaier and Wambach, 2010; Goel and Thakor, 2006; Grund and Sliwka, 2005; Itoh, 2004; Rey-Biel, 2007). When workers dislike pay inequality, team incentives may outperform both individual and relative incentives, because team incentives generate little inequality of pay among workers. We differ from these studies in that workers are altruistic rather than inequity-averse, and that workers' altruism is endogenously determined. One implication is that – in line

with the empirical evidence mentioned at the end of the previous paragraph – the introduction of team-based incentives on top of flat wages increases the quality of co-worker relations in our model, while it is neutral in models of inequity aversion. Moreover, our results on the effects of relative incentives are also clearly different from those that arise when workers are inequity averse.

2.3 The model

We consider a profit-maximizing principal who employs two homogenous agents.[4] Agents produce output by exerting effort. Effort of agent i is denoted by $e_i \geq 0$. Total profits of the principal are:

$$\pi = Q(e_i, e_j) - w_i - w_j,$$

where the production function Q satisfies the Inada conditions with respect to all inputs, and w_i denotes agent i's wage.

Agents care about three things: their wage, their cost of effort, and their net benefit from social interaction with colleagues. We model social interaction as an exchange of attention between agents. We assume that receiving attention is pleasurable, while giving attention is costly.[5] The utility function of agent i is:

$$U_i = w_i - C(e_i, a_{ij}) + G(a_{ji}) + F(a_{ji}, U_j), \tag{2.1}$$

where $a_{ij} \geq 0$ denotes the attention given by agent i to agent j. The cost function C is strictly convex and increasing in both arguments and satisfies the conditions $C(0,0) = 0$, $C_e(0, a_{ij}) = 0$, and $C_a(e_i, 0) = 0$, where subscripts to functions denote partial derivatives. For simplicity, we assume that $C_{ea}(\cdot) = 0$.[6] Receiving attention

[4] Our results generalize to the case of an arbitrary number $n > 2$ of agents. Details can be found in the appendix.

[5] These assumptions are stronger than we need: They only need to hold at the margin in the optimum. For instance, allowing agents to enjoy giving attention up to some point would not change our results qualitatively. Clearly, in practice, receipt of attention is sometimes costly; in those cases, our results do not carry over.

[6] Clearly, giving and receiving attention takes time and, hence, may increase worker's marginal cost of effort. On the other hand, as shown by some of the studies we discussed in the introduction,

generates two types of benefits to an agent, represented by the functions G and F. First, we allow for a consumption benefit from attention, captured by the strictly concave and increasing function G, with $G_a(0) = +\infty$.[7] Second, we assume that receipt of attention leads to feelings of altruism for the giving agent, which increases the utility of the receiving agent. This is captured by the function $F(a_{ji}, U_j) = \gamma a_{ji} U_j$, where $\gamma > 0$. The specific functional form keeps the analysis tractable.[8] To ensure an interior solution, we shall abstract from situations where $F_u \geq 1$.[9] That is, agents always care more for themselves than for others. Last, note that the linearity of utility in income implies that the agent is risk-neutral.

The principal offers a contract to each agent that makes each agent at least as well off as his outside option $\overline{U} > 0$. The principal can condition the agent's wage on the effort of the agent himself and also on the effort of his colleague ($w^i(e_i, e_j)$). We shall speak of individual incentives when $w^i_{e_i}(e_i, e_j) > 0$, of team incentives when $w^i_{e_j}(e_i, e_j) > 0$, and of relative incentives when $w^i_{e_j}(e_i, e_j) < 0$.

The timing of the game is as follows. In the first stage, the principal offers contracts to the agents, which they accept or reject. Next, agents decide simultaneously and independently how much attention to give to their co-worker. In the last stage, agents decide on the level of effort they exert.[10]

---

social contact with colleagues can reduce stress and increase job involvement. In practice, both of these arguments are likely to play a role, which may explain why the empirical evidence on the relation between worker cohesiveness and productivity is rather mixed (see Rotemberg 2006 for an overview). Itoh (1991b) provides a thorough analysis of the consequences of these kind of interdependencies between effort and attention provision in a closely related principal-multi-agent model.

[7] As will become clear, our results are qualitatively the same in the absence of a consumption benefit from receiving attention.

[8] Clearly, the specific functional form implies that receipt of attention increases the utility of the receiving agent only when the giving agent has positive utility, which we assume. This assumption is, however, inessential. The analysis would be unchanged if the function F depended on the difference between the utility that the giving agent obtains and the utility U^s that he would obtain if the receiving agent provided the 'selfish' levels of effort and attention (that is, $F(a_{ji}, U_j) = \gamma a_{ji} (U_j - U^s)$).

[9] Clearly, this is ensured when agents do not get a positive utility from $a \geq \frac{1}{\gamma}$. Note that it is easy to extend the function to allow for unconditional altruism or spite, e.g. $F(a_{ji}, U_j) = \gamma(a_{ji} + \phi)U_j$, where $\phi \neq 0$. This would not affect our results qualitatively except for situations where unconditional spite is very strong so that creation of co-worker altruism through social interaction is inefficient.

[10] Obviously, in the one-shot game that we study, the exact timing is important for the results. In a repeated setting where agents alternately give attention and provide effort, this would be much less of a concern. In such a setting, it is important that the agents have a finite horizon

2.4 Complete contract

We start by studying the benchmark case where both effort and attention are contractible. Full contractibility implies that there is no reason to condition the wage on effort, and so an agent's compensation only consists of a base salary in this section. The principal's optimization problem is:

$$\max_{e_i,e_j,a_{ij},a_{ji},w_i,w_j} Q(e_i,e_j) - w_i - w_j$$

subject to the agents' participation constraints:

$$w_i - C(e_i,a_{ij}) + G(a_{ji}) + F(a_{ji},U_j) \geq \overline{U}, \tag{2.2}$$

$$w_j - C(e_j,a_{ji}) + G(a_{ij}) + F(a_{ij},U_i) \geq \overline{U}. \tag{2.3}$$

The first-best levels of effort and attention are described in Proposition 2.1.

Proposition 2.1 *The complete contract has the following properties:*

1. *Effort of each agent is strictly positive and equates the marginal benefits of effort to the principal with the marginal cost of effort to the agent:* $Q_e(\cdot) = C_e(\cdot)$*;*
2. *Attention by each agent is strictly positive and equates the receiving agent's marginal benefits with the giving agent's marginal cost of attention:* $G_a(\cdot) + F_a(\cdot) = C_a(\cdot)$*.*
3. *The wage makes each agent indifferent between accepting and rejecting the contract, given the first-best levels of effort and attention:* $w = \overline{U} + C(\cdot) - G(\cdot) - F(\cdot)$*.*

The proof is given in the appendix. As usual, the first-best contract induces agents to exert the level of effort that maximizes the joint surplus. The principal optimally includes a positive level of attention in the contract. Even though attention

(e.g., retirement). In an infinite horizon model, first-best attention may arise as equilibrium play without any contractual arrangements by the principal.

entails a cost for the giving agent, it produces a pleasant working environment for the receiving agent, which allows the principal to pay a lower wage.

2.5 Incomplete contracts

Next let us consider the more realistic case where workers' attention is not contractible; the principal cannot contract on workers' actions like showing interest in a colleague's personal life, treating colleagues with courtesy, or giving affective support. We keep the assumption of contractible effort.[11] As we shall see, the principal finds it optimal to condition each agent's wage on the effort of both agents, $w(e_i, e_j)$. For convenience, we assume that the wage contract is linear in both e_i and e_j. This is innocuous: As will become clear, the principal can not do better by offering nonlinear contracts. Let $w^i_{e_i}$ denote agent i's bonus per unit of effort provided by agent i (representing individual incentives) and let $w^i_{e_j}$ denote agent i's bonus per unit of effort provided by agent j (representing team or relative incentives). Further, let s^i be agent i's base salary. We solve the maximization problem of the principal by backward induction, starting with the agent's choice of effort.

The first-order condition for agent i's optimal effort is described by:

$$w^i_{e_i}(\cdot) + w^j_{e_i}(\cdot)F_{u_j}(\cdot) - C_{e_i}(\cdot) = 0. \tag{2.4}$$

Effort has three effects on an agent's utility. First, when the principal gives individual incentives, the agent's wage increases with his effort. Second, when the principal has installed team incentives or relative incentives, agent i's effort choice affects agent j's income. Agent i cares about this effect to the extent that he is altruistic towards his colleague. Last, there is a cost of providing effort. The optimal effort level equates these benefits and costs at the margin.

The comparative static effect of social interaction on the agent's effort is summarized in the following Lemma.

[11] None of the results change when effort is noncontractible as long as the principal can contract on a (noisy) signal of each agent's effort (e.g. output). Extending the model to allow for risk aversion of agents in the presence of noisy signals of effort does not affect our results qualitatively.

Lemma 2.1 *Social interaction affects the agent's choice of effort as follows:*

1. *The effect of received attention on effort is described by:*

$$\frac{de_i}{da_{ji}} = \frac{F_{u_j a_{ji}}(\cdot) w^j_{e_i}(\cdot)}{C_{e_i e_i}(\cdot)}, \tag{2.5}$$

   implying that an agent's effort increases with received attention when the contract includes team incentives, while effort decreases with received attention when the contract includes relative incentives.

2. *Attention given by agent i has no effect on his effort:*

$$\frac{de_i}{da_{ij}} = 0.$$

The first part of Lemma 2.1 echoes the results by Rotemberg (1994) and Bandiera et al. (2005) on the relation between co-worker altruism and effort. When workers care for one another, they partly take into account the externalities they impose on others. Compared to egoistic agents, this implies higher effort under team incentives and lower effort under relative incentives. As co-worker altruism increases with received attention, effort increases with attention under team incentives and it decreases with attention under relative incentives. The second part of Lemma 2.1 directly follows from the separability of effort cost and attention cost in the worker's utility function. Clearly, when effort and attention would be substitutes, agent's effort would decrease with attention given by the agent.

In the previous stage of the game, the agents decide independently on how much attention to give to their co-worker, taking into account the effect of their attention on effort in the next stage of the game. The first-order condition for agent's optimal attention is:

$$\frac{dU_i}{da_{ij}} = -C_{a_{ij}}(\cdot) + \frac{dU_j}{da_{ij}} F_{u_j}(\cdot) + \frac{de_i}{da_{ij}} \frac{dU_i}{de_i} + \frac{de_j}{da_{ij}} \frac{dU_i}{de_j} = 0. \tag{2.6}$$

Using the agents' first-order conditions for optimal effort (2.4), this can be simplified

to:

$$-C_{a_{ij}}(\cdot) + \frac{dU_j}{da_{ij}} F_{u_j}(\cdot) + \frac{de_j}{da_{ij}} \frac{dU_i}{de_j} = 0. \tag{2.7}$$

Besides the direct cost of attention provision, giving attention has two effects on the agent's utility. First, when the agent has altruistic feelings towards his co-worker, he enjoys the increase in his co-worker's utility resulting from the receipt of attention. Differentiating (2.1), it follows that the increase in the co-worker's utility is:

$$\frac{dU_j}{da_{ij}} = G_{a_{ij}}(\cdot) + F_{a_{ij}}(\cdot) + F_{u_i}(\cdot) \frac{dU_i}{da_{ij}}, \tag{2.8}$$

where the last term drops by the envelope theorem, using (2.6). Second, there is an indirect effect of attention provision through the co-worker's effort choice: By giving more attention, the agent induces the co-worker to change his level of effort in the next stage, which in turn affects the agent's utility. Differentiating (2.1) it follows that:

$$\frac{dU_i}{de_j} = w^i_{e_j}(\cdot) + F_{u_j}(\cdot) \frac{dU_j}{de_j} = w^i_{e_j}(\cdot), \tag{2.9}$$

where the last equality follows from applying the envelope theorem, using the first-order condition for optimal effort (2.4). Clearly, an agent's utility is only affected by his co-worker's effort when the contract includes team incentives or relative performance incentives. Similarly, we learned from Lemma 2.1 that a worker's effort is only affected by received attention when the contract has team or relative incentives. Taking these two results together, it follows that the last term of the first-order condition (2.7) is strictly positive when either team incentives or relative incentives are part of the agent's contract. That is, both team incentives and relative incentives create an additional marginal benefit from attention provision for each agent. This benefit stems from the effect of attention-giving on co-worker altruism and, hence, on effort. When the contract has team incentives, an agent's provision of attention induces the other agent to exert more effort in the next stage, which benefits the agent. Likewise, when the contract has relative incentives, the agent's provision of attention induces the other agent to exert less effort in the next stage, which again benefits the agent. Lemma 2.2 follows.

Lemma 2.2 *Both team incentives and relative incentives promote social interaction among workers.*

Intuitively, when team incentives are provided, agents invest in altruism so as to foster cooperation. When relative incentives are provided, agents invest in altruism so as to tame their colleague's eagerness to outperform.

Substituting (2.5), (2.8) and (2.9) into (2.7) gives agent i's first-order condition for optimal attention in a rewritten form:

$$-C_{a_{ij}}(\cdot) + F_{u_j}(\cdot)\left[G_{a_{ij}}(\cdot) + F_{a_{ij}}(\cdot)\right] + \frac{F_{u_i a_{ij}}(\cdot) w^i_{e_j}(\cdot)}{C_{e_j e_j}(\cdot)} w^i_{e_j}(\cdot) = 0. \tag{2.10}$$

Our next result follows immediately and is described in the following Proposition.

Proposition 2.2 *When the principal does not provide team incentives or relative incentives ($w^i_{e_j} = 0$ for all $i \neq j$), there is too little social interaction in equilibrium.*

Proposition 2.2 follows from a comparison of the first-order condition for attention (2.10) with first-best attention as described by Proposition 2.1. In the absence of team or relative incentives, the last term of first-order condition (2.10) drops. Comparing with the first-best as described by Proposition 2.1, it follows that there is too little social interaction in any equilibrium where $F_u < 1$, as we have imposed. That is: As agents care less about their co-worker than about themselves, the benefits from attention provision are not fully internalized. Underprovision of attention results. Note that there exist multiple equilibria. First, an equilibrium exists where neither of the agents give attention. When an agent believes that the other agent will not give any attention, the second term of (2.10) is zero, implying that the agent only faces costs from attention provision (as reflected by the first term of (2.10)). Hence, given that an agent expects to receive no attention, it is optimal for him to give no attention as well. Second, depending on the exact shape of the functions, one or more equilibria with positive attention exist. When the function $G(\cdot)$, representing the consumption benefits from attention, is sufficiently concave, or the

cost function $C(\cdot)$ is sufficiently convex in attention, there is a unique equilibrium with strictly positive attention.[12] Anyway, since in all possible equilibria attention is described by (2.10), attention is always at a suboptimal level. As a result, the principal's profits are lower than first-best.[13]

Last, consider the principal's contract design problem, which is given by:

$$\max_{w^i_{e_i},w^i_{e_j},w^j_{e_j},w^j_{e_i},s^i,s^j} Q(e_i,e_j) - (w^i_{e_i} + w^j_{e_i})e_i - (w^i_{e_j} + w^j_{e_j})e_j - s^i - s^j,$$

subject to the agents' participation constraints (2.2) and (2.3). First-best effort, attention, and profits are achieved by the incentive contract described in Proposition 2.3.

Proposition 2.3 *When attention cannot be contracted, but effort can, the principal achieves first-best profits by offering an incentive contract consisting of a base salary, individual incentives, and team or relative incentives. Optimal individual incentives are described by:*

$$w^i_{e_i} = Q_{e_i}(\cdot) - F_{u_j}(\cdot)w^j_{e_i} \quad \textit{for } i \neq j,$$

and optimal team or relative incentives are described by:

$$w^i_{e_j} = \pm\sqrt{\frac{(1 - F_{u_j}(\cdot))\left[G_{a_i}(\cdot) + F_{a_i}(\cdot)\right]C_{ee}(\cdot)}{F_{u_i a_{ij}}(\cdot)}} \quad \textit{for } i \neq j,$$

where all functions are evaluated at the first-best levels of effort and attention, as described by Proposition 2.1. The level of the base salary follows from the agents' participation constraints.

The proof is in the appendix. The principal can obtain maximum profits by including a mix of individual incentive pay and team or relative incentive pay in the

[12] The appendix describes the precise condition; it rules out that agent's responsiveness to received attention increases with received attention. In the remainder of this chapter, we shall assume that this condition holds.

[13] Note that, since this result holds for any level of the cost of attention, companies that restrict their workplace policies to facilitating social interaction (that is, reducing agent's marginal cost of attention) will achieve lower than first-best profits.

contract. As we have seen in Lemma 2.2, team or relative incentives can be used to promote social interaction. In the optimum, the team incentives or relative incentives are chosen such that, given first-best effort, the agents provide first-best attention as described in Proposition 2.1. Next, the principal can ensure that agents exert first-best effort by adjusting the individual incentives. When attention provision is stimulated through team incentives, individual incentives are muted, since an agent enjoys the positive effect his effort has on his colleague's wage. With relative incentives, the effort of an agent negatively influences the utility of his co-worker, which in equilibrium is an additional cost of effort. Individual incentives therefore need to be adjusted upwards to restore efficient effort provision.

2.6 Concluding remarks

For many employees, social interaction with colleagues is one of the key determinants of job satisfaction. We have studied the influence of financial incentives for productive activities on the quality of co-worker relationships in a model where agents have endogenous other-regarding preferences. Following earlier work on the formation of social ties, we have assumed that the strength of a worker's altruistic feelings towards a colleague is increasing with the colleague's kindness towards the worker. We have seen that, absent team or relative incentives, workers do not invest enough in their relationships with their co-workers, as the benefits from relationship-building to the colleague are not fully internalized. This externality problem comes at a cost to the employer, as good co-worker relationships allow employers to attract and retain workers without paying high wages. We have shown that employers can stimulate social interaction among colleagues by providing either team incentives or relative incentives.

We have deliberately kept the analysis as simple as possible. An interesting next step would be to study situations where the employer can only contract on team output, so that team incentives serve a dual role: promoting productive effort and stimulating social interaction. In such situations, too much social interaction may arise (as strong team incentives may be optimal to boost production, but as a side-

effect create too much concern among workers to please each other). In response to this, employers may take actions so as to increase workers' cost of giving attention. Likewise, employers may not be so keen on encouraging social interaction when relative performance incentives are the sole instrument to promote effort. Other interesting extensions include heterogeneity in workers' social preferences and the sorting of different types of workers to firms offering different incentive schemes. Kosfeld and von Siemens (2009, 2011) make interesting steps in that direction.

2.A Appendix

Proof of Proposition 2.1

Let λ_i denote the Lagrange-multiplier for the participation constraint of agent i. The first-order conditions are:

$$Q_{e_i}(\cdot) + \lambda_i \frac{dU_i}{de_i} + \lambda_j \frac{dU_j}{de_i} = 0; \qquad Q_{e_j}(\cdot) + \lambda_i \frac{dU_i}{de_j} + \lambda_j \frac{dU_j}{de_j} = 0; \tag{2.A1}$$

$$\lambda_i \frac{dU_i}{da_{ij}} + \lambda_j \frac{dU_j}{da_{ij}} = 0; \qquad \lambda_i \frac{dU_i}{da_{ji}} + \lambda_j \frac{dU_j}{da_{ji}} = 0; \tag{2.A2}$$

$$-1 + \lambda_i \frac{dU_i}{dw_i} + \lambda_j \frac{dU_j}{dw_i} = 0; \qquad -1 + \lambda_i \frac{dU_i}{dw_j} + \lambda_j \frac{dU_j}{dw_j} = 0; \tag{2.A3}$$

and the two participation constraints. Using (2.A1) and (2.A3) and noting that

$$\begin{aligned} \frac{dU_j}{de_i} &= F_u(\cdot)\frac{dU_i}{de_i}, \\ \frac{dU_j}{dw_i} &= F_u(\cdot)\frac{dU_i}{dw_i}, \\ \frac{dU_i}{de_i}\left(\frac{dU_i}{dw_i}\right)^{-1} &= -C_e(\cdot), \end{aligned}$$

the first part of Proposition 2.1 follows. Substituting

$$\begin{aligned} \frac{dU_i}{da_{ij}} &= -C_a(\cdot) + F_u(\cdot)\frac{dU_j}{da_{ij}} \\ \frac{dU_j}{da_{ij}} &= G_a(\cdot) + F_a(\cdot) + F_u(\cdot)\frac{dU_i}{da_{ij}} \end{aligned}$$

into (2.A2) yields after successive substitution:

$$\left[-C_a(\cdot) + G_a(\cdot) + F_a(\cdot)\right]\left[1 + F_u(\cdot) + F_u^2(\cdot) + ... + F_u^{\infty}(\cdot)\right] = 0,$$

from which the second part of Proposition 2.1 follows. The third part follows from (2.A1) and (2.A3) which imply that $\lambda_i = \lambda_j > 0$, and hence the participation constraints bind.

Condition for unique equilibrium with strictly positive attention

Agent i's best-response curve has the following slope:

$$\frac{da_{ij}}{da_{ji}} = -\frac{d^2U_i/da_{ij}da_{ji}}{d^2U_i/da_{ij}^2} = -\frac{F_{u_j a_{ji}}(\cdot)\left[G_{a_{ij}}(\cdot) + F_{a_{ij}}(\cdot)\right] + F_{u_j}(\cdot)F_{a_{ij}u_i}(\cdot)\frac{dU_i}{da_{ji}}}{-C_{a_{ij}a_{ij}}(\cdot) + F_{u_j}(\cdot)G_{a_{ij}a_{ij}}(\cdot)}, \quad \text{(2.A4)}$$

$$\text{where } \frac{dU_i}{da_{ji}} = G_{a_{ji}}(\cdot) + F_{a_{ji}}(\cdot) + F_{u_j}\frac{dU_j}{da_{ji}}. \quad \text{(2.A5)}$$

Note that at the origin the slope of the best-response curve is infinitely large, as $G_{a_{ij}}(0) \to +\infty$. Hence, to ensure that there is a unique equilibrium with strictly positive attention, it is sufficient that the best-response curve is strictly concave when $da_{ij}/da_{ji} > 0$. The second derivative of the best-response curve is given by the following expression:

$$\frac{d^2a_{ij}}{(da_{ji})^2} = -\frac{(d^3U_i/da_{ij}da_{ji}^2)(d^2U_i/da_{ij}^2) - (d^3U_i/da_{ij}^2da_{ji})(d^2U_i/da_{ij}da_{ji})}{\left(d^2U_i/da_{ij}^2\right)^2},$$

where d^2U_i/da_{ij}^2 and $d^3U_i/da_{ij}^2da_{ji}$ are always negative by the second-order condition and by the concavity of the $G(\cdot)$ function, respectively, and $d^2U_i/da_{ij}da_{ji}$ is always positive in the relevant area where $da_{ij}/da_{ji} > 0$ (see (2.A4)). Hence, a sufficient (but not necessary) condition for the best-response function to be strictly concave when $da_{ij}/da_{ji} > 0$ is:

$$\begin{aligned} &d^3U_i/da_{ij}da_{ji}^2 = 2\gamma^2\left[-C_{a_{ji}}(\cdot) + \gamma a_{ij}\left(G_{a_{ji}}(\cdot) + \gamma U_i\right)\right] + \\ &\gamma^2 a_{ji}\left\{-C_{a_{ji}a_{ji}}(\cdot) + \gamma a_{ij}\left[G_{a_{ji}a_{ji}}(\cdot) + \gamma\left(dU_i/da_{ji}\right)\right]\right\} < 0 \end{aligned} \quad \text{(2.A6)}$$

which is satisfied when the $G(\cdot)$ function is sufficiently concave or the $C(\cdot)$ function is sufficiently convex in attention.

Proof of Proposition 2.3

The agent's choice of attention is described by first-order condition (2.10). Comparing with Proposition 2.1, to achieve first-best attention, the principal should set incentives such that the last two terms of (2.10) equal $G_a(\cdot) + F_a(\cdot)$, or:

$$F_{u_j}(\cdot)\left[G_{a_{ij}}(\cdot) + F_{a_{ij}}(\cdot)\right] + \frac{F_{u_i a_{ij}}(\cdot) w^i_{e_j}}{C_{e_j e_j}(\cdot)} w^i_{e_j} = G_a(\cdot) + F_a(\cdot).$$

Solving for $w^i_{e_j}$ gives the expression for relative or team incentives in Proposition 2.3. The optimal level of individual incentives follows along similar lines, using first-order condition (2.4) and the expression for first-best effort in Proposition 2.1. Lastly, note that we do not need to be concerned about multiplicity of equilibria since, first, the equilibrium where both workers abstain from giving attention is no longer an equilibrium when $w^i_{e_j} \neq 0$, and, second, there exists only one equilibrium with strictly positive attention when condition (2.A6) holds, which is assumed.

Generalization to n-agents

The analysis in the main text can be generalized to an arbitrary number of identical agents, where the proof follows the same lines as the proofs for 2 agents. The profits of the principal are:

$$\pi = Q(e_1, ..., e_n) - \sum_i w_i,$$

and agent i's utility function is:

$$U_i = w_i - C(e_i, \sum_{j \neq i} a_{ij}) + \sum_{j \neq i} G(a_{ji}) + \sum_{j \neq i} F(a_{ji}, U_j).$$

In this generalization of the results we proceed as follows, first, we have to set a boundary on the level of altruism. Second, we derive the first-best effort, attention, and wage levels similar to Proposition 2.1. Finally, we demonstrate that with unobservable attention individual incentives alone do not lead to the first-best, while it can be obtained using team incentives and relative performance incentives, parallel to Propositions 2.2 and 2.3.

For the analysis with n-agents we need to impose a stricter condition on the level

of altruism. Our analysis with two agents illustrated a sort of 'social multiplier', as agent i not only cared for agent j's well-being, but also that agent j cares for i, and that agent j cared that i cares about his well-being, etcetera , which converges only if $F_u < 1$. With a larger number of agents convergence is more difficult, because a third or fourth agent benefits from the good relationships between his colleagues. That is, agent k is not only altruistic towards agent i and agent j, but he also enjoys that agent i and agent j are altruistic towards him and each other. To ensure an interior solution the level of altruism needs to satisfy $F_u < \frac{1}{n-1}$, which we derive next.[14]

After successive substitution the utility function becomes:

$$U_i = w_i - C(\cdot) + \sum_{j \neq i} G(\cdot) + \sum_{m=2}^{n} \left[(m-1) \frac{n!}{m!(n-m)!} F_u^m \right] U_i. \tag{2.A7}$$

From (2.A7) we can see the need for the assumption:

$$\sum_{m=2}^{n} \left[(m-1) \frac{n!}{m!(n-m)!} F_u^m \right] < 1. \tag{2.A8}$$

[14]Like in the main text, we actually need to assume that agents do not choose attention levels where $F_u \geq \frac{1}{n-1}$, which is ensured if $a \geq \frac{1}{\gamma(n-1)}$ leads to a negative utility.

We can rewrite the expression from (2.A8):

$$\begin{aligned}
&\sum_{m=2}^{n}\left[(m-1)\frac{n!}{m!(n-m)!}F_u^m\right] \\
&= F_u^2\sum_{m=2}^{n}\left[(m-1)\frac{n!}{m!(n-m)!}F_u^{m-2}\right] \\
&= F_u^2\sum_{m=2}^{n}\left[\frac{n!}{m!(n-m)!}\frac{d}{dF_u}F_u^{m-1}\right] \\
&= F_u^2\frac{d}{dF_u}F_u^{-1}\sum_{m=2}^{n}\left[\frac{n!}{m!(n-m)!}F_u^m\right] \\
&= F_u^2\frac{d}{dF_u}F_u^{-1}\sum_{m=0}^{n}\left[\left(\frac{n!}{m!(n-m)!}F_u^m\right)-\frac{n!}{0!(n-0)!}F_u^0-\frac{n!}{1!(n-1)!}F_u^1\right] \\
&= F_u^2\frac{d}{dF_u}F_u^{-1}\left[(1+F_u)^n-1-nF_u\right] = F_u^2\left[\frac{n(1+F_u)^{n-1}F_u-\left[(1+F_u)^n-1\right]}{F_u^2}\right] \\
&= 1+n(1+F_u)^{n-1}F_u-(1+F_u)^n = 1-(1+F_u)^{n-1}\left[1-(n-1)F_u\right],
\end{aligned}$$

which allows us to simplify (2.A8):

$$\begin{aligned}
\sum_{m=2}^{n}\left[(m-1)\frac{n!}{m!(n-m)!}F_u^m\right] < 1 &\Leftrightarrow 1-(1+F_u)^{n-1}\left[1-(n-1)F_u\right]<1 \Leftrightarrow \\
&\Leftrightarrow (1+F_u)^{n-1}\left[1-(n-1)F_u\right]>0 \Leftrightarrow \\
&\Leftrightarrow 1-(n-1)F_u>0 \Leftrightarrow F_u<\frac{1}{(n-1)},
\end{aligned}$$

where the last line follows as $F_u > -1$, which was already satisfied the assumptions $y > 0$ and $a \geq 0$.

To derive the first-best levels of effort and attention we look at the principal who maximizes profits subject to the n participation constraints ($U_i = w_i - C(e_i, a_i) + \sum_{j\neq i} G(a_{ji}) + \sum_{j\neq i} F(a_{ji}, U_j) \geq U_0$). This constrained optimization problem has the following first-order conditions:

$$Q_{ei}(\cdot) + \lambda_i \frac{dU_i}{de_i} + \sum_{j \neq i} \lambda_j \frac{dU_j}{de_i} = 0; \text{ for } \forall i, i \neq j \tag{2.A9}$$

$$-1 + \lambda_i \frac{dU_i}{dw_i} + \sum_{j \neq i} \lambda_j \frac{dU_j}{dw_i} = 0; \text{ for } \forall i, i \neq j \tag{2.A10}$$

$$\lambda_i \frac{dU_i}{da_{ij}} + \lambda_j \frac{dU_j}{da_{ij}} + \sum_{k \neq i,j} \lambda_k \frac{dU_k}{da_{ij}} = 0; \text{ for } \forall i \text{ and } \forall j, i \neq j \neq k \tag{2.A11}$$

and the participation constraints. As

$$\frac{dU_j}{de_i} = F^j_{u_i}(\cdot)\frac{dU_i}{de_i} + \sum_{k \neq i,j} F^j_{u_k}(\cdot)\frac{dU_k}{de_i},$$

and $\frac{dU_j}{de_i} = \frac{dU_k}{de_i}$, since the colleagues are affected similarly by agent i's effort, we can rewrite (2.A9):

$$Q_{ei} + \frac{dU_i}{de_i}\left[\lambda_i + \sum_{j \neq i} \lambda_j \left[\frac{F^j_{u_i}(\cdot)}{1 - \sum_{k \neq i,j} F^j_{u_k}(\cdot)}\right]\right] = 0.$$

Similar, (2.A10) can be written as:

$$-1 + \frac{dU_i}{dw_i}\left[\lambda_i + \sum_{j \neq i} \lambda_j \left[\frac{F^j_{u_i}(\cdot)}{1 - \sum_{k \neq i,j} F^j_{u_k}(\cdot)}\right]\right] = 0.$$

Using

$$\frac{dU_i}{de_i}\left(\frac{dU_i}{dw_i}\right)^{-1} = -C_{ei}(\cdot),$$

from which the first part of Proposition 2.1 follows. Likewise, noting that

$$\frac{dU_k}{da_{ij}} = F^k_{u_i}(\cdot)\frac{dU_i}{da_{ij}} + F^k_{u_j}(\cdot)\frac{dU_j}{da_{ij}} + \sum_{l \neq i,j,k} F^k_{u_l}(\cdot)\frac{dU_l}{da_{ij}}, \tag{2.A12}$$

and $\frac{dU_k}{da_{ij}} = \frac{dU_l}{da_{ij}}$, as the attention from agent i to agent j affects other (third party)

colleagues the same, we can rewrite (2.A11):

$$\frac{dU_i}{da_{ij}}\left[\lambda_i + \sum_{k\neq i,j}^{n} \lambda_k \left[\frac{F_{u_i}(\cdot)}{1-\sum_{l\neq i,j,k} F_{u_l}(\cdot)}\right]\right] +$$

$$\frac{dU_j}{da_{ij}}\left[\lambda_j + \sum_{k\neq i,j} \lambda_k \left[\frac{F_{u_j}(\cdot)}{1-\sum_{l\neq i,j,k} F_{u_l}(\cdot)}\right]\right] = 0. \tag{2.A13}$$

Given that agents are identical, we know that $\lambda_i = \lambda_j$, (2.A13) simplifies to $\frac{dU_i}{da_{ij}} + \frac{dU_j}{da_{ij}} = 0$. (2.A12) and $\frac{dU_k}{da_{ij}} = \frac{dU_l}{da_{ij}}$ can be used to express $\frac{dU_i}{da_{ij}}$ in $\frac{dU_j}{da_{ij}}$ and vice versa, which allows for a similar successive substitution argument as in the end of the proof of Proposition 2.1. This gives the second part of Proposition 2.1 and completes the generalization for the first-best.

The generalization of Propositions 2.2 and 2.3, where the agent's attention choice is unobservable follows the same steps as in the main text, solving the game backwards. In the derivation we assume that all individuals receive a similar linear incentive scheme of the form: $w = w^i_{e_i} e_i + \sum_{j\neq i} w^i_{e_j} e_j$, however this innocuous as it holds true for profit maximizing incentives. Further, as we will only look at a symmetric equilibrium, all F_u's may be set equal.

Agent i's first-order condition for optimal effort is:

$$\frac{dU_i}{de_i} = w^i_{ei}(\cdot) - C_{ei}(\cdot) + \sum_{j\neq i} F^i_{u_j}(\cdot)\frac{w^j_{ei}}{1-\sum_{k\neq j,i} F^j_{u_k}} = 0 \tag{2.A14}$$

where there last term follows from:

$$\frac{dU_j}{de_i} = w^j_{ei} + F^j_{u_i}(\cdot)\frac{dU_i}{de_i} + \sum_{k\neq j,i} F^j_{u_k}(\cdot)\frac{dU_k}{de_i} = \frac{w^j_{ei}}{1-\sum_{k\neq j,i} F^j_{u_k}}, \tag{2.A15}$$

combined with the envelope theorem. Note that, since agent i's effort level has a similar influence on the wage of each of his colleagues, we may put $\frac{dU_j}{de_i} = \frac{dU_k}{de_i}$. From the agent's first-order condition (2.A14) we can derive comparative statics

with respect to received and provided attention:

$$\frac{de_i}{da_{ij}} = 0, \tag{2.A16}$$

$$\frac{de_i}{da_{ji}} = \frac{F_{u_j a_{ji}} w^j_{e_i}}{C_{e_i e_i}(1 - \sum_{k \neq j,i} F^j_{u_k})} \tag{2.A17}$$

$$\frac{de_i}{da_{kj}} = \frac{F_{u_k a_{kj}} F_{u_j} w^j_{e_i}}{C_{e_i e_i}(1 - \sum_{k \neq j,i} F^j_{u_k})^2}. \tag{2.A18}$$

Agent i's first-order condition for giving attention to agent j is as follows:

$$\frac{dU_i}{da_{ij}} = -C_{aij}(\cdot) + F^i_{u_j}(\cdot)\frac{dU_j}{da_{ij}} + \sum_{k \neq j,i} F^i_{u_k}(\cdot)\frac{dU_k}{da_{ij}} + \frac{de_j}{da_{ij}}\frac{dU_i}{de_j} + \sum_{k \neq i,j} \frac{de_k}{da_{ij}}\frac{dU_i}{de_k} = 0. \tag{2.A19}$$

The attention from agent i to agent j affects the other agents in the following way:

$$\begin{aligned} \frac{dU_j}{da_{ij}} &= G_{a_{ij}} + F_{a_{ij}} + \sum_{k \neq j,i} F^j_{u_k}(\cdot)\frac{dU_k}{da_{ij}} + \sum_{k \neq i,j} \frac{de_k}{da_{ij}}\frac{dU_j}{de_k} \\ &\quad + F^j_{u_i}(\cdot)\frac{dU_i}{da_{ij}} + \frac{de_i}{da_{ij}}\frac{dU_j}{de_i}, \end{aligned} \tag{2.A20}$$

$$\begin{aligned} \frac{dU_k}{da_{ij}} &= F^k_{u_j}(\cdot)\frac{dU_j}{da_{ij}} + \frac{de_j}{da_{ij}}\frac{dU_k}{de_j} + \sum_{l \neq k,j,i} F^k_{u_l}(\cdot)\frac{dU_l}{da_{ij}} + \sum_{l \neq k,j,i} \frac{de_l}{da_{ij}}\frac{dU_k}{de_l} \\ &\quad + F^k_{u_i}(\cdot)\frac{dU_i}{da_{ij}} + \frac{de_i}{da_{ij}}\frac{dU_k}{de_i}. \end{aligned} \tag{2.A21}$$

The expressions (2.A20) and (2.A21) need to be rewritten to substitute them into (2.A19), because they contain each other. First, note that in both (2.A20) and (2.A21) the last two terms drop, due to the envelope theorem and (2.A16). Further, since we look at a symmetric equilibrium, we can save on notation by combining (2.A17) and (2.A18), which gives $\frac{de_k}{da_{ij}} = \frac{de_l}{da_{ij}} = \frac{F_u}{1-(n-2)F_u}\frac{de_j}{da_{ij}}$ and use $\frac{dU_i}{de_j} = \frac{dU_j}{de_i}$ for $\forall i$, $\forall j$, and $i \neq j$ as well. In addition, we may replace the summations $\sum_{k \neq j,i}$ and $\sum_{l \neq k,j,i}$ with $(n-2)$ and $(n-3)$ respectively. Finally, noticing that $\frac{dU_k}{da_{ij}} = \frac{dU_l}{da_{ij}}$, one

can substitute a slightly rewritten (2.A21) into (2.A20), which yields:

$$\frac{dU_j}{da_{ij}} = (G_{a_{ij}} + F_{a_{ij}})\frac{1-(n-3)F_u}{1-(n-3)F_u-(n-2)F_u^2} + \frac{de_j}{da_{ij}}\frac{dU_i}{de_j}\frac{(n-2)F_u\left[2-(n-2)F_u\right]}{\left[1-(n-3)F_u-(n-2)F_u^2\right]\left[1-(n-2)F_u\right]}, \tag{2.A22}$$

substituting (2.A22) back into (2.A21) gives:

$$\frac{dU_k}{da_{ij}} = (G_{a_{ij}} + F_{a_{ij}})\frac{F_u}{1-(n-3)F_u-(n-2)F_u^2} + \frac{de_j}{da_{ij}}\frac{dU_i}{de_j}\frac{1-F_u+(n-2)F_u^2}{\left[1-(n-3)F_u-(n-2)F_u^2\right]\left[1-(n-2)F_u\right]}. \tag{2.A23}$$

The combination of (2.A19) with (2.A22) and (2.A23) presents:

$$\frac{dU_i}{da_{ij}} = -C_{aij}(\cdot) + (G_{a_{ij}} + F_{a_{ij}})\frac{F_u\left[1-(n-3)F_u\right]+(n-2)F_u^2}{1-(n-3)F_u-(n-2)F_u^2} + \frac{de_j}{da_{ij}}\frac{dU_i}{de_j}\frac{1-(n-3)F_u+(n-2)F_u}{\left[1-(n-3)F_u-(n-2)F_u^2\right]\left[1-(n-2)F_u\right]}. \tag{2.A24}$$

One can verify, using $F_u < \frac{1}{n-1}$, that agents will provide attention too little without team incentives or relative performance incentives, as the last term of (2.A24) drops in this case. Intuitively, by having $F_u < \frac{1}{n-1}$, we ensured that there is still an internalization problem in the provision of incentives. To derive the profit maximizing team incentives or relative performance incentives, we substitute (2.A15) and (2.A17) into (2.A24), which gives:

$$w^i_{e_j} = \pm\sqrt{\frac{\left[G_{a_{ij}}(\cdot)+F_{a_{ij}}(\cdot)\right]C_{e_je_j}(\cdot)}{F_{u_ia_{ij}}(\cdot)}\frac{\left[1-(n-2)F_u-(n-1)F_u^2\right]\left[1-(n-2)F_u\right]}{1+F_u}} \quad \text{for } i \neq j,$$

when set equal to the first-best level of attention. The remainder of the generalization follows from solving (2A.14) for first-best effort and setting base salaries such that participation constraints bind.

Chapter 3

Peer Evaluation: Incentives and Co-worker Relations

3.1 Introduction

In many employment relations managers lack information about employees' individual effort or performance; e.g. in highly interdependent work teams, with work on location, or in case of experts. In an attempt to obtain a more complete picture of employee performance, many firms have turned to multisource feedback, in business better known as 360° evaluations. A 360° evaluation can include performance assessments by subordinates, peers, supervisors, customers, or other stakeholders. According to survey data, about 90% of Fortune 1000 firms use some form of multisource feedback, often including evaluation by peers (Edwards and Ewen 1996). Moreover, in many companies peer evaluation schemes partially determine personnel decisions regarding promotions and performance pay (Bohl 1996).[1]

Evaluation by peers has intuitive appeal when employees have valuable information about one another's performance. However, when seen from an economic theory perspective, the widespread use of peer evaluation is less obvious. A major concern is that employees may benefit from providing the employer with invalid performance

[1] Bohl (1996) finds that one-fifth of 750 surveyed companies reports to use peer evaluations, of which 90% includes the peer evaluation in personnel decisions. Similarly, in a survey among 280 firms, roughly one-fifth of the firms have adopted some form of peer evaluation (Antonioni 1996).

evaluations of their colleagues. For instance, employees can provide invalid ratings so as to help their friends, hurt their competitors, or 'game the system' through collusion (Edwards and Ewen 1996 and Kozlowski et al. 1998). In line with this, Tsui and Barry (1986) find that performance evaluations are more lenient when raters experience more personal affect towards the ratee.[2] Love (1981) even asserts that: "The most common rationale given by organizational personnel for avoiding the use of peer assessment is that it is simply a "popularity contest." " (p. 451). This chapter develops a formal agency model so as to study when and how peer evaluation should be used under different intensities of co-worker relations.

I study peer evaluation in a model of team production by two homogeneous agents, who are protected by limited liability. The production suffers from a pure moral hazard problem, i.e., the principal is unable to distinguish individual inputs, and can only offer a team bonus for output. A team bonus results in inefficiently low effort levels, since limited liability makes the budget breaking solution infeasible. Agents receive a signal about their colleague's effort provision. In an attempt to offer more efficient incentives for effort, the principal employs peer evaluation to gather the information from these signals. A bonus is paid to those agents that receive a positve evaluation report. As a starting point I assume ideal circumstances for these peer evaluations: Agents have an aversion towards lying about their signal. A small cost of lying is sufficient to ensure truthful evaluations, as the agents' reports only affect the payoff of their colleague. Peer evaluation motivates agents to exert effort so as to increase the likelihood that their co-worker receives a positive signal and rates performance accordingly. Und er these ideal circumstances, peer evaluation outperforms a team bonus; in fact, peer evaluation performs as well as individual incentives would.

Next, I allow for interpersonal relations between colleagues, either good or bad. Social preferences can give agents an incentive to lie about their signal, i.e., result in a (dis)likeability bias. In case the internalized utility from a bonus to a friend outweighs the cost of lying, agents in good relationships give positive evaluations

[2]Similar significant correlations between likeability or friendships and leniency in peer evaluation can be found in Sonnentag (1998) and Love (1981). Antonioni and Park (2001) show that peer evaluations suffer more from such a likeability bias than traditional performance evaluations.

regardless of their signal. Likewise, agents in bad relationships may begrudge their colleague a bonus and lie upon receiving a positive signal. Invalid peer evaluations also provide incentives for effort, through internalization of a colleague's lying costs. Profits, however, are usually higher when peer evaluations remain truthful. The principal can ensure truthful evaluation by adjusting the bonus for peer evaluations downward, such that co-worker relations do not affect the evaluation decision. Social preferences, thereby, constrain the effectiveness of peer evaluation. Nevertheless, the optimal contract always includes peer evaluation. The principal will complement incentives for effort by including a team bonus in case peer evaluation becomes severely constrained.

The combination of peer evaluation and a team bonus in the optimal contract gives an interesting comparative static result regarding how co-worker relations affect profits. The effectiveness of the team bonus increases with better co-worker relations, as co-worker relations mitigate incentives to free-ride on a colleague. Peer evaluation suffers from more pronounced co-worker relations, through the earlier discussed likability bias. Taken together, co-worker relations have non-monotonic effects on profits.

Despite the widespread use of peer evaluation, there are few empirical studies on its effectiveness. The success of peer evaluation programs is difficult to determine, since objective performance measures are often lacking, motivating peer assessment to start with. Peiperl (1999) is one of the few empirical studies that looks into success factors of peer evaluation in business.[3] Managers in a multinational service-sector organization were asked to identify units with (un)successful peer evaluation. Successful peer evaluation was positively correlated with a high interdependency of work within and across units, group components in incentives, the perceived quality of the ratings, and its integration with rewards. In addition, peer evaluation was found to be less successful in units that had a very positive group culture, of which Peiperl says: "an unexpected result that bears some consideration" (p. 445). The likability bias illustrated in this chapter can explain this observation.

[3]Bamberger et al. (2005) evaluate the impact of a peer assessment program in a kibbutz-owned manufacturing facility in Israel. The peer evaluations led to an improvement of performance (measured by supervisory ratings), especially so in the nonanonymous departments.

The results of this chapter, that better co-worker relations erode peer evaluation, while they strengthen the effectiveness of a team bonus, are largely in line with findings by Towry (2003). In a laboratory experiment, she studies how the effectiveness of these two incentive schemes are influenced by a team identity manipulation. In this experiment, participants either receive a compensation based on a colleague's report about performance, i.e. peer evaluation, or they receive team bonus. Towry finds that the peer evaluations are undermined by strong team identity, because team members became more likely to cover for each other's shirking behavior. The team incentive, on the other hand, was more effective in this strong team identity treatment.

The chapter is organized as follows. The next section gives a brief overview of the related literature. Section 3.3 describes the set-up of the model, followed by the analysis in Section 3.4. Section 3.5 contains concluding remarks.

3.2 Related literature

In a pioneering paper, Ma (1988) shows that there exist mechanisms where it is in the agents' best interest to report shared information truthfully. Ma considers a principal who is unable to observe the inputs of a team of agents, while agents observe each others' actions perfectly. The principal can implement first-best effort by asking one agent to report the effort choices of all team members, and giving other agents the option to challenge this report. There exists a lottery for which the option to challenge a report is exercised if and only if the report contains a lie. Since other agents will always challenge a false report, it is optimal for the reporting agent to state the chosen actions truthfully. The principal, thereby, retrieves the agents' information on individual inputs, which enables him to implement first-best effort.

This chapter differs from the set-up by Ma in the agents' ability to observe each others' action: Instead of perfect observability, or an assumption that signals are sufficient statistics for effort (Fischer and Hughes 1997), I assume that agents get a coarse signal of a colleague's performance. In the set-up I consider, the principal does not benefit from asking agents to report their own effort choice in addition to

the signal about a co-worker's effort. The coarse signals do not allow colleagues to verify the stated reports perfectly, hence the principal cannot play off agents' reports against each other. As a consequence, agents have no incentive to report their effort truthfully, rendering the mechanism above infeasible. In many work settings, e.g., within health care, consultancy, or academics, the performance assessment by peers is relevant, yet it is rarely based on perfect information about effort.

Instead of interpersonal relations, Barron and Gupte (2009) focus on other limitations of peer evaluation. The authors introduce harmful consequences of negative peer appraisals, modelled in two ways: First, negative appraisals make agents less productive through a drop in team cohesion. Second, agents experience a loss of utility due to peer pressure. Barron and Gupte show that, in a Ma-type of lottery for truthtelling, agents need higher compensation to report shirking by a colleague.[4] Peer evaluation is no longer guaranteed to be profitable under these adverse effects of a negative appraisal, instead profitability depends on the quality of signals among peers.

Marx and Squintani (2009) also analyze a pure moral hazard problem in team production, where agents can be given the task to evaluate each other. Without peer evaluation, the combination of unobservable individual effort choices, and an inability to punish individuals without proof of shirking behavior, results in low output. The principal can achieve the first-best outcome with a contract that demands both high effort and peer monitoring, notwithstanding that monitoring is costly. The peer monitoring gives an incentive to provide effort, as the principal can now punish upon observing low output. Punishment is possible, since peer monitoring provides verifiable evidence of shirking, or in case evidence is lacking, agents failed to monitor. The principal will not punish agents for not monitoring with high output, because he cannot distinguish between non-monitoring and the absence of evidence for shirking. Therefore, the agents exert effort and refrain from monitoring in equilibrium, leading to the first-best. The authors show that their results are sensitive to the monitoring technology.

[4]The lottery offered in Barron and Gupte (2009) plays off the signal that a co-worker receives against a signal the principal has, while Ma (1988) only uses the information that the agents share.

3.3 The model

I consider a principal who employs two homogeneous agents, denoted by i and j, and maximizes the following profit function:

$$\pi = qQ_H - \sum_i \left[w_{Li} + q\Delta w + m_{ji} B_i \right], \text{ where } i \neq j. \tag{3.1}$$

The principal's profit equals output minus wages. Output can be high or low, where high output has value Q_H and low output is set to zero. The output realization is determined by random variable $q \in \{0, 1\}$, where the probability of high output is increasing in the effort of both agents: $\Pr(q = 1) = \varphi(e_i, e_j) = e_i + e_j$.[5] The principal only observes the output level that is realized, i.e., effort choices are unobservable. Wages can contain a base wage, w_{Li}, a team bonus for high output, Δw, and a bonus for the receipt of a positive peer evaluation, B_i.[6] Wages are subject to a limited-liability constraint: The wage cannot be negative in any state of the world. This gives the following limited-liability constraints:

$$w_{Li} \geq 0, \; w_{Li} + B_i \geq 0, \; w_{Li} + \Delta w \geq 0, \; w_{Li} + B_i + \Delta w \geq 0. \tag{3.2}$$

Agents maximize the following utility function:

$$U_i = w_{Li} + q\Delta w + m_{ji} B_i - C(e_i) - \gamma \left| m_{ij} - s_{ij} \right| + \alpha U_j. \tag{3.3}$$

Utility is increasing in the wage and decreasing with costs of effort. For convenience, I assume quadratic costs of effort: $C(e_i) = \frac{1}{2} e_i^2$. The principal asks agents to evaluate the performance of their colleague: Agent i sends an evaluation message to

[5] The assumption of a probabilistic team output is not essential, it is important that individual inputs are unobservable to the principal. It does simplify the analysis by the reduction of the number of possible team bonuses. Further, it creates the need for a restriction on Q_H, or on the cost of effort, such that high output with certainty is not preffered over the optimal contract.

[6] Hereby, I restrict my attention to a limited class of contracts. I consider the performance of the least sophisticated form of peer evaluation incentives, where the peer evaluation bonus only depends on the report by a colleague. This may not be without loss of generality: In a more general setting, the principal could specify a contract $w_i(m_{ij}, m_{ji}, q)$, which would allow for 8 possible bonus payments in the binary setting that I analyse. The optimal contract for more sophisticated peer evaluation schemes, i.e., where payments depend on the interaction of reports and team outcomes, is left for further research.

the principal about agent j's performance, denoted by m_{ij}. Agents can base their evaluation on a signal they receive about their colleague's effort level. The signal, $s_{ij} \in \{0, 1\}$, takes on zero when agent i perceives j's effort is below some standard and one if j performed well according to i. The probability of a high signal depends on the colleague's effort, $\Pr(s_{ij} = 1) = p(e_j) = e_j$. I assume that there are some costs of deviating from one's signal, or $\gamma > 0$. These costs can stem from the time and energy spent to misrepresent information (Kartik 2009). Besides this, many people bear a psychological costs of lying (Gneezy 2005, Hurkens and Kartik 2009, Fischbacher and Heusi 2008). The last term in the utility function captures social preferences in the workplace. A positive α captures altruism, whereas a negative α represents feelings of spite. As a consequence of the homogeneity assumption, α can also be interpreted as the quality of a relationship between colleagues. Substitution of U_j into (3.3) gives:

$$U_i = \frac{1}{1-\alpha^2} \left\{ \begin{array}{c} w_{Li} + q\Delta w + m_{ji}B_i - C(e_i) - \gamma \left| m_{ij} - s_{ij} \right| \\ +\alpha \left[w_{Lj} + q\Delta w + m_{ij}B_j - C(e_j) - \gamma \left| m_{ji} - s_{ji} \right| \right] \end{array} \right\},$$

where α is restricted to $-1 < \alpha < 1$, to ensure an interior solution.[7]

The timing of the game is as follows: First, there is a contracting stage, where the agents can accept or reject the offered contract. Next, each agent chooses effort. Finally, the agents receive a signal about their co-worker's effort level and send an evaluation message to the principal, after which output is realized and the contracts are executed.

3.4 Analysis

In a situation without peer evaluation, the principal can only use the team bonus to stimulate effort provision. The team bonus suffers from a well-known internalization

[7] The term $\frac{1}{1-\alpha^2}$ is like a social multiplier of utility. In case of positive relations, agent i not only enjoys the utility of j, but also that j values i's well-being. In case of bad relations, j's utility enters as a cost in U_i. However, i's fortune also makes j less well off, so that i's 'private' utility is again amplified. Still, if agents are identical, the latter effect is second-order, and agents are worse off with bad co-worker relations. At the end of section 3.4, I will discuss comparative statics with respect to α.

problem; as the agents' costs of effort are private, while the benefits are not fully internalized, the free-rider problem arises. Still, the principal can achieve first-best profits with a team bonus when budget breaking is possible (Holmstrom, 1982). The principal pays the full marginal product to both agents, and achieves first-best profits through a negative base wage. This solution is not feasible here, as agents are assumed to be limitedly liable. Hence, a team bonus cannot implement first-best effort. Next, let us see whether it is possible to improve upon the team bonus by peer evaluations.

The game is solved backwards, starting with the peer evaluation stage. As a consequence of the binary signal, the principal believes at best two messages in equilibrium, $m_{ij} \in \{0, 1\}$.[8] Agents choose the evaluation message that maximizes their utility. A low evaluation brings a cost of lying in case the received signal was high. Similarly, giving a high evaluation brings a cost of lying when the signal was low, and it leads to a bonus for the colleague. The agent's evaluation remains truthful ($m_{ij} = s_{ij}$), as long as $m_{ij} = 0$ is preferred for $s_{ij} = 0$:

$$U_i(m_{ij} = 0) - U_i(m_{ij} = 1) \geq 0, \text{ which holds for } \gamma - \alpha B_j \geq 0,$$

and $m_{ij} = 1$ is preferred with $s_{ij} = 1$:

$$U_i(m_{ij} = 1) - U_i(m_{ij} = 0) \geq 0, \text{ which holds for } \gamma + \alpha B_j \geq 0.$$

Obviously, without social preferences (i.e., $\alpha = 0$), the assumption of lying costs leads to a truthful revelation of one's signal. However, once we allow for co-worker relations, agents may have an incentive to lie. Namely, the bonus for a positive evaluation gives an internalized benefit or cost depending on the agent's social preferences. Agents in good relationships may want to lie when receiving a low signal about their colleague's performance: In case the utility of a bonus for a friend outweighs the lying costs, agents give each other positive evaluations irrespective of the received signal. Likewise, colleagues in a bad relationship may want to lie in case of

[8] For other messages, I assume that the principal believes the agent received a low signal, and awards no bonus.

a positive signal, as they begrudge the bonus to a disliked colleague. Truthful evaluation is thereby constrained to $-\frac{\gamma}{B} \leq \alpha \leq \frac{\gamma}{B}$; i.e., when agents have weak social preferences, high costs of lying, or there is a small bonus for a positive evaluation. Lemma 3.1 summarizes an agent's evaluation strategy:

Lemma 3.1 *Agent i's evaluation strategy is:* $m_{ij} = \begin{cases} 0 & \text{if } \alpha < -\frac{\gamma}{B_j} \\ s_{ij} & \text{if } -\frac{\gamma}{B_j} \leq \alpha \leq \frac{\gamma}{B_j} \\ 1 & \text{if } \alpha > \frac{\gamma}{B_j} \end{cases}$.

The evaluation process is truthful when agents have weak social preferences, high lying costs, or small benefits from a positive evaluation. This translates into a constraint on the bonus for the receipt of a positive peer evaluation under which peer evaluations remain truthful:

$$B \leq \frac{\gamma}{|\alpha|}. \tag{3.4}$$

In the preceding stage agents choose their effort levels. The first-order condition for effort is the following:

$$(1-\alpha^2)\frac{dU_i}{de_i} = (1+\alpha)\varphi_{e_i}(\cdot)\Delta w - C_{e_i}(\cdot) + \frac{dU_i}{d\Pr(s_{ji}=1)}\frac{d\Pr(s_{ji}=1)}{de_i} = 0.$$

Effort provision increases with the team bonus for high output, while it decreases with the marginal costs of effort. The team bonus also affects the income of a colleague. The latter effect can be an additional benefit or costs of effort depending on agent i's social preferences. The team bonus provides stronger incentives when agent i is altruistic towards j, while effort provision is suppressed when i is spiteful towards j. The last term in the first-order condition captures incentives for effort from peer evaluation. The expectation that a co-worker receives a positive signal is increasing in effort, $\frac{d\Pr(s_{ji}=1)}{de_i} = p_{e_i}(\cdot) = 1$. A change in the co-worker's signal affects utility as follows:

$$\frac{dU_i}{d\Pr(s_{ji}=1)} = \begin{cases} -\alpha\gamma & \text{if } m_{ij} = 0 \\ B_i & \text{if } m_{ij} = s_{ij} \\ \alpha\gamma & \text{if } m_{ij} = 1 \end{cases}. \tag{3.5}$$

Combined, these terms give two incentive effects. First, in case of truthful peer evaluation, effort is stimulated by the peer evaluation bonus. Agents desire to increase the likelihood that their colleague receives a positive signal and rates performance accordingly, such that the bonus is obtained. Second, under non-truthful peer evaluation there is also an incentive for effort, despite that the signal about a co-worker's performance is ignored in the evaluation decision. Altruistic agents wish to increase the probability that their colleague receives a positive signal, such that the expected lying costs for this colleague are lower. Likewise, spiteful agents can increase the lying costs of a colleague by increasing the probability of a positive signal. The incentives from non-truthful peer evaluation are weaker than those of the maximum bonus under truthful peer evaluation:

$$B = \frac{\gamma}{|\alpha|} > |\alpha|\,\gamma \text{ as } |\alpha| < 1.$$

For non-truthful peer evaluation the bonus for receiving a positive evaluation does not give incentives for effort itself, however the bonus should be sufficiently large to give rise to non-truthful peer evaluation strategies.[9] It is evident that non-truthful peer evaluations is unattractive to the principal when co-worker relations are good, as the incentives are weaker and the higher B always needs to be paid. In case of bad co-worker relations, the non-truthful equilibrium has potential to be more profitable, as the peer evaluation bonus is never paid. Nonetheless, the principal often achieves higher profits with truthful peer evaluation, for which a sufficient condition is:

$$\alpha\gamma > -\frac{1}{2}(1+\alpha)Q_H, \tag{3.6}$$

see the appendix for the proof. In case this condition does not hold, the principal may want to exploit a bad relationship between agents by setting a high peer evaluation bonus, which is never paid. As described before, agents have an incentive to work, because they can increase the expected lying costs of a disliked colleague. In the

[9]Note that, peer evaluation without a bonus, or $B_i = 0$, gives no incentive for effort. Developmental peer evaluation, or peer evaluation without compensation, is a much used tool, but without value in the setting I study. Recently, Gupte (2009) presents a model of developmental peer evaluation, where peer evaluation is used as an input to determine a suitable training.

remainder of this chapter I will assume that (3.6) holds, since the practical relevance of this non-truthful equilibrium is to my opinion limited. Consequently, we can focus on the optimal effort choice under truthful peer evaluation. In combination with the assumptions on $\varphi(\cdot)$ and $C(\cdot)$, the first-order condition for effort implies that agent i's optimal effort under truthful peer evaluation equals:

$$e_i = (1+\alpha)\Delta w + B_{i.} \tag{3.7}$$

Summarizing the above: Effort can be stimulated through a team bonus and a bonus for receiving a positive peer evaluation. The effectiveness of the team bonus depends on an agent's social preferences. The free-rider problem, inherent to a team bonus, is less severe when co-worker relations are good. The free-rider problem is exacerbated for bad co-worker relations, as agents begrudge their colleague the fruits of their labour. Further, a bonus for the receipt of a positive peer evaluation stimulates effort provision, because effort influences a colleague's expected evaluation message under truthful peer evaluation.

Finally, the principal sets the profit maximizing contracts subject to the incentive compatibility constraints (3.7), the limited-liability constraints (3.2), and the constraint for truthful peer evaluation (3.4). Naturally, the principal does not need to abide to this last constraint (3.4). The inclusion of this constraint helps to characterize the optimal contract, but is innocuous, as the principal cannot do better with non-truthful peer evaluation when (3.6) holds. Together, this gives the principal's optimization problem:

$$\begin{aligned} \max_{w_{Li},\Delta w,B_i,B_j} \pi &= \varphi(\cdot)\left[Q_H - 2\Delta w\right] - w_{Li} - w_{Lj} - E(m_{ij})B_j - E(m_{ji})B_i, \\ \text{s.t. } w_{Li} &\geq 0,\ w_{Li} + B_i \geq 0,\ w_{Li} + \Delta w \geq 0,\ w_{Li} + B_i + \Delta w \geq 0, B_i \leq \frac{\gamma}{|\alpha|}, \end{aligned}$$

The first limited-liability constraint will bind, as there is no incentive for effort from the base wage. Hence, we get $w_{Li} = 0$. The last limited-liability constraint is satisfied by a combination of the former constraints. This reduces the optimization

problem to:

$$\begin{aligned} \max_{\Delta w, B_i, B_j} \pi &= \varphi(\cdot)\left[Q_H - 2\Delta w\right] - w_{Li} - w_{Lj} - E(m_{ij})B_j - E(m_{ji})B_i, \quad (3.8) \\ \text{s.t. } \Delta w &\geq 0,\ B_i \geq 0, B_i \leq \frac{\gamma}{|\alpha|}, \end{aligned}$$

leading to the second-best contract described in Proposition 3.1:

Proposition 3.1 *Under assumption (3.6), the optimal contract under limited liability is characterized by:*

$$B_i = B_j = \frac{1}{2}Q_H \text{ and } \Delta w = 0, \text{ if } \gamma \geq \frac{1}{2}|\alpha| Q_H, \quad (A)$$

$$B_i = B_j = \frac{\gamma}{|\alpha|} \text{ and } \Delta w = 0, \text{ if } \frac{(1+\alpha)|\alpha| Q_H}{3+\alpha} \leq \gamma \leq \frac{1}{2}|\alpha| Q_H, \quad (B)$$

$$B_i = B_j = \frac{\gamma}{|\alpha|} \text{ and } \Delta w = \frac{Q_H}{4} - \frac{(3+\alpha)\gamma}{4(1+\alpha)|a|}, \text{ if } \gamma < \frac{(1+\alpha)|\alpha| Q_H}{3+\alpha}. \quad (C)$$

The proof can be found in the appendix. The optimal contract always includes a bonus for receiving a positive peer evaluation. The principal can ensure peer evaluation remains truthful, by satisfying constraint (3.4). This allows the principal to offer individual incentives for effort through B_i, even though he does not observe the individual contributions himself. Hence, unlike the team bonus, the peer evaluation bonus does not suffer from an internalization problem, which makes it a more cost-effective way to stimulate effort.

Part (A) of Proposition 3.1 shows the optimal contract under ideal circumstances for peer evaluation, i.e. sufficiently high cost of lying or weak social preferences. In this situation, the peer evaluation bonus performs as well as providing the agents with individual incentives under limited liability. Smaller lying costs or more pronounced social preferences constrain the peer evaluation bonus at some point, described by contract (B). Eventually, the principal will find it optimal to complement incentives for effort with a team bonus, as indicated by contract (C) in Proposition 3.1.

The comparative statics of α and γ with respect to the peer evaluation bonus

and the team bonus are described by Corollary 3.1:

Corollary 3.1 *In contracts (B) and (C), the bonus for the receipt of a positive peer evaluation is increasing in lying costs and social preferences that become less pronounced:*

$$\begin{aligned} \frac{\partial B}{\partial \gamma} &= \frac{1}{|\alpha|} > 0, \\ \frac{\partial B}{\partial \alpha} &= -\frac{\alpha\gamma}{|\alpha|^3}. \end{aligned}$$

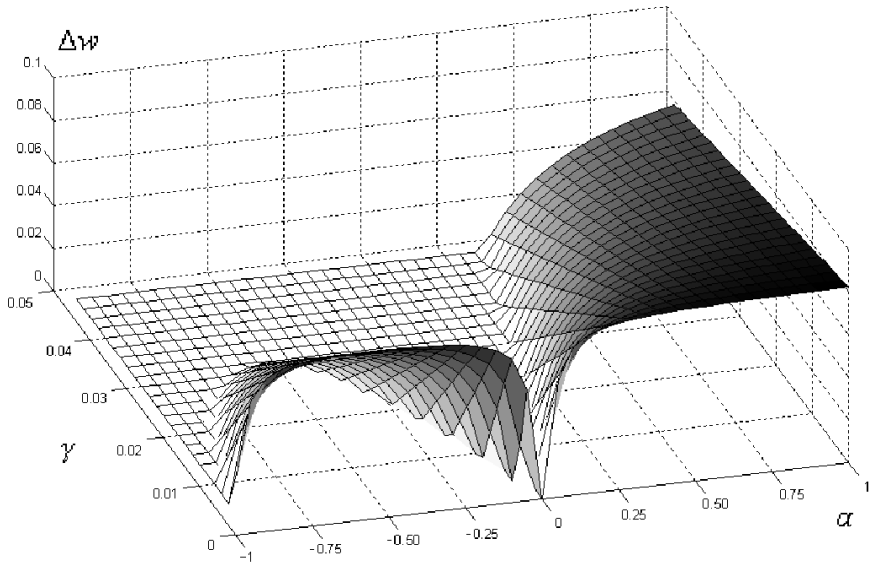
In contract (C), the team bonus is decreasing with the costs of lying, and social preferences have an ambiguous effect on the team bonus:

$$\begin{aligned} \frac{\partial \Delta w}{\partial \gamma} &= -\frac{\partial B}{\partial \gamma}\frac{(3+\alpha)}{4(1+\alpha)} < 0, \\ \frac{\partial \Delta w}{\partial \alpha} &= -\frac{\partial B}{\partial \alpha}\frac{(3+\alpha)}{4(1+\alpha)} + \frac{1}{2(1+\alpha)^2}\frac{\gamma}{|\alpha|}. \end{aligned} \tag{3.9}$$

Otherwise, lying costs and social preferences do not affect compensation in the second-best contract.

The comparative static result with respect to the peer evaluation bonus follows from Lemma 3.1. That is, decreasing costs of lying and more pronounced co-worker relations constrain the peer evaluation bonus. Corollary 3.1 shows a negative relation between lying costs and the team bonus. A decrease in the costs of lying paralyzes peer evaluation, making it optimal to enhance effort provision with a higher team bonus (provided that lying costs are sufficiently low). Social preferences have a twofold influence on the strength of the team bonus, as can be seen from (3.9). First, like with lying costs, stronger social preferences constrain the bonus for receiving a positive peer evaluation. More pronounced co-worker relations, whether negative or positive, thereby add to the importance of the team bonus. Second, social preferences also influence the incentive effects of a team bonus, as we already observed in (3.7). For altruistic agents better co-worker relations strengthen the team bonus by both effects. The two effects are in conflict for bad co-worker relations that become worse,

Figure 3.1: Team bonus in the optimal contract for different (γ, α) combinations, with $Q_H = \frac{1}{3}$.



which leads to a decrease of the team bonus at some point for more spiteful agents. The optimal team bonus in the second-best contract is depicted in Figure 3.1 for different (γ, α) combinations, with $Q_H = \frac{1}{3}$. Figure 3.2 and Figure 3.3 do so for the expected profits and expected utility.[10] The comparative statics for the expected profits and expected utility can be found in the appendix.

As long as peer evaluation is unrestricted (A), the expected profits are constant. In contract (B), lower lying costs and more pronounced co-worker relations hurt profits, through the smaller incentive effect of the peer evaluations. Finally, in contract (C), the size of the peer evaluation bonus and the team bonus are both influenced by γ and α. A decrease of the lying costs leads to a shift in the compensation towards the team bonus. This shift hurts profits, as the peer evaluation bonus stimulates effort in a more cost-effective way than the team bonus does, so profits are still increasing in γ under contract (C). There is a similar effect of more pronounced co-worker relations, a shift away from the peer evaluation bonus has a negative effect on profits.

[10] The figures are qualitatively the same for other levels of Q_H. The sufficient condition for optimality of truthful peer evaluation (3.6) does not hold for some (γ, α) combinations in the upper left corners in Figure 3.1 to 3.3, in which case the optimal contract may look different.

Figure 3.2: Expected profits under the optimal contract for different (γ, α) combinations, with $Q_H = \frac{1}{3}$.

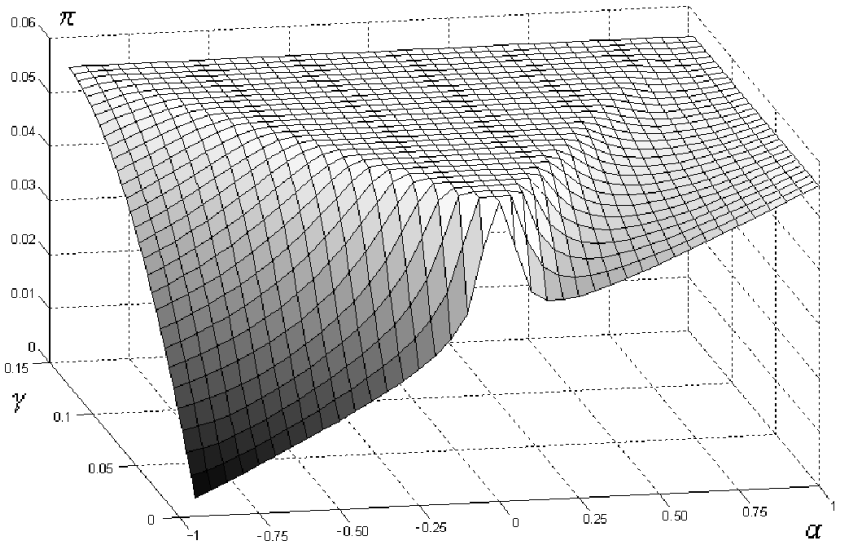
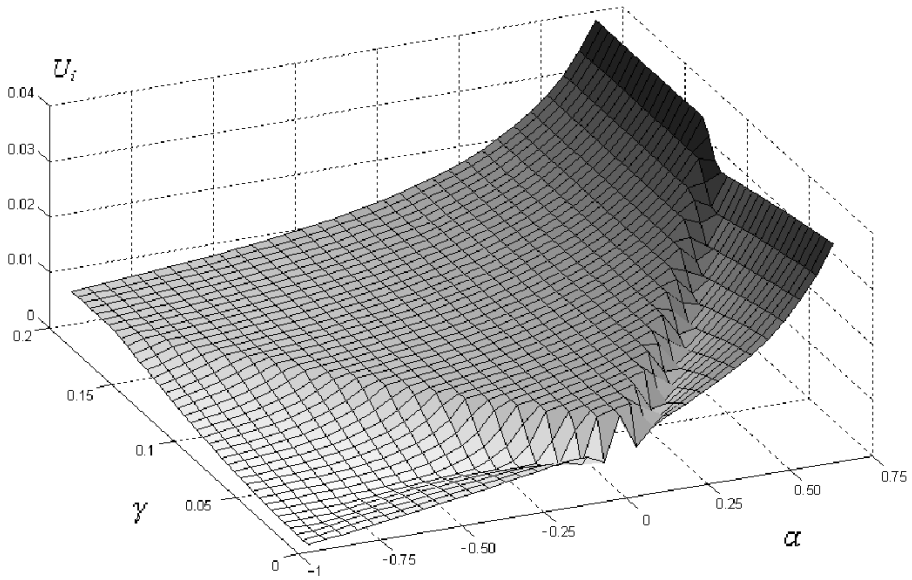


Figure 3.3: Expected utility in the optimal contract for different (γ, α) combinations, with $Q_H = \frac{1}{3}$.



In addition, better co-worker relations soften the free-rider problem, which has a favorable effect on profits. Profits are increasing in α with bad co-worker relations, as both effects work in the same direction. For good co-worker relations, profits are decreasing in α at first, as the shift away from peer evaluation dominates the effect of a softened free-rider problem. However, the profits pick up again from some point, as the effect on the free-rider problem becomes more prominent for a larger team bonus. Summarizing, under truthful peer evaluation, bad co-worker relations that become worse are never good for profits, while improving good co-worker relations can be bad.

Agents are left with rents due to the limited liability assumption. Lying costs only affect utility through the offered contract, since there is no lying in equilibrium. Therefore, utility is constant in γ for unrestricted peer evaluation (A). Co-worker relations, on the other hand, have a direct and positive effect on utility, as already noted in footnote 7. In contract (B), an increase in the lying costs relaxes the constraint on the peer evaluation bonus, and thereby leads to a higher utility. Similarly, besides the direct effect on utility, co-worker relations affect the size of the peer evaluation bonus. The expected utility is increasing in α, except for moderately positive co-worker relations ($0 < \alpha < \frac{2}{3}$), where the decrease in the peer evaluation bonus outweighs the direct effect of having better co-worker relations. Finally, in contract (C), higher lying costs cause a shift in the optimal compensation towards peer evaluation, as noticed before. This shift away from the team bonus hurts the agents' utility, as the team bonus comes with higher rents. A change in α also has this effect, in addition to the direct effect on utility and the effect on the free-rider problem. The latter two effects make the utility increasing in co-worker relations, whereas the first effect is negative for $\alpha < 0$ and can dominate for moderately negative co-worker relations.

3.5 Concluding remarks

Peer evaluation has intuitive appeal in work environments where co-workers are in the best position to monitor each other. However, there is a concern that these

performance evaluations by peers are susceptible to a likeability bias. I have studied the possibility to include peer evaluation in performance pay under different intensities of co-worker relations, using a model of team production by two limitedly liable agents. The principal only observes the team output, while agents receive a signal about a colleague's effort, allowing for peer evaluation in addition to a team bonus. The combination of a cost of lying about the received signal and interpersonal relations between co-workers determine the agent's evaluation decision. Lying costs lead to a truthful revelation of signals, while the internalization of a bonus to a colleague may cause a bias. Nevertheless, the optimal contract always includes a bonus for receiving a positive peer evaluation, which may be complemented with a team bonus. Truthful peer evaluation has incentive effects, as agents desire to increase the likelihood that their colleague receives a positive signal. Social preferences constrain the bonus for a positive evaluation so as to keep revelation of signals truthful, leading to the testable implication that incentive effects of peer evaluation diminish with more pronounced social relations between co-workers. A managerial implication of this result is that managers should assign employees that are rather indifferent towards one another to evaluate each other's performance.

There is debate in the management literature and organizational psychology whether peer evaluations should only be used for developmental purposes, or whether they can also be used to determine merit pay and promotions (see among others Edwards and Ewen 1996, Coates 1998, Fleenor and Brutus 2001). I contribute to this debate by showing that some concerns of intergrating peer evaluation with performance pay can be overcome. Namely, by setting a smaller bonus for peer evaluation, it is possible to avoid a (dis)likability bias. In addition, I illustrated that even non-truthful peer evaluations can have incentive effects. The distinction between developmental peer evaluation and its integration with rewards, however, may be less strict in practice: In case a manager is supposed to serve as a coach, it is unlikely that the information resulting from peer evaluations is neglected in determining promotions and pay (Edwards and Ewen 1996). The consequences of including peer evaluation in performance pay, therefore, deserve more research attention.

There are a number of other factors that may lead to intentional distortions in peer evaluation. Peer evaluation between colleagues in promotion tournaments will most likely lead to a downward bias in the evaluation reports, in order to boost one's own chances of winning. Carpenter et al. (2010) provide evidence for this subtle way of sabotaging competitors in a real effort experiment, where the anticipation of sabotage eliminates the incentives to work in this tournament. A special case of peer evaluation is where incumbents determine their future co-workers, or competitors for future promotions, as incumbents may be most capable in evaluating the quality of applicants. Carmichael (1988) showed that in such a setting incumbents discriminate against the best possible candidates, unless the incumbents' future income is ensured, which is sometimes only possible by offering tenure. Another concern is that colleagues may try to collude with their evaluations against the principal in repeated interactions.

The result that a principal always wants to make use of peer evaluations is at odds with its occurrence in practice. However, the prerequisite that co-workers have valuable information on employee performance does not apply to all organizations. Further, it is conceivable that peer evaluation has an influence on interpersonal relationships. Chapter 2 showed that, under the assumption of conditionally altruistic or reciprocal preferences, externalities in wages may motivate people to invest in their relationships with colleagues. In line with this, peer evaluation may be a good reason to invest in the relationship with a colleague that evaluates one's performance, so as to increase the likelihood of a positive evaluation by this colleague. By this logic the effectiveness of peer evaluation may deteriorate eventually. In combination with some of the additional concerns named above, this offers an explanation for the gap between the theoretical prediction and the prevalence of peer evaluations in practice.

3.A Appendix

Proof of Proposition 1

The agents' effort choices (3.7) are substituted into $\varphi(\cdot)$ and $E(m_{ij}) = p(\cdot)$, which gives the following Kuhn-Tucker conditions to the principal's optimization problem (3.8):

$$\begin{aligned}
&\frac{d\pi}{d\Delta w} = 2(1+\alpha)Q_H - 8(1+\alpha)\Delta w - (3+\alpha)B_i - (3+\alpha)B_j + \lambda_1 \leq 0,\\
&\Delta w \geq 0,\ \Delta w\frac{d\pi}{d\Delta w} = 0,\ \lambda_1 \geq 0,\ \lambda_1\Delta w = 0,\\
&\frac{d\pi}{dB_i} = Q_H - (3+\alpha)\Delta w - 2B_i + \lambda_2 - \lambda_3 \leq 0,\ B_i \geq 0,\ B_i\frac{d\pi}{dB_i} = 0,\\
&\lambda_2 \geq 0,\ \lambda_2 B_i = 0,\ B_i - \frac{\gamma}{|\alpha|} \leq 0,\ \lambda_3 \geq 0,\ \lambda_3\left[B_i - \frac{\gamma}{|\alpha|}\right] = 0,\\
&\frac{d\pi}{dB_j} = Q_H - (3+\alpha)\Delta w - 2B_j + \lambda_4 - \lambda_5 \leq 0,\ B_j \geq 0,\ B_j\frac{d\pi}{dB_j} = 0,\\
&\lambda_4 \geq 0,\ \lambda_4 B_i = 0,\ B_j - \frac{\gamma}{|\alpha|} \leq 0,\ \lambda_5 \geq 0,\ \lambda_5\left[B_j - \frac{\gamma}{|\alpha|}\right] = 0.
\end{aligned}$$

We look for a solution where $B_i = B_j$, as agents are homogeneous. Further, as the constraints on the peer evaluation bonus can never bind at the same time, the cases that need be checked are reduced to six: (i) none of the constraints is binding, (ii) $\Delta w = 0$, and $B_i = B_j = 0$, (iii) $\Delta w > 0$ and $B_i = B_j = 0$, (iv) $\Delta w = 0$ and constraints on $B_i = B_j$ are non-binding, (v) $\Delta w > 0$ and $B_i = B_j = \frac{\gamma}{|\alpha|}$, and (vi) $\Delta w = 0$ and $B_i = B_j = \frac{\gamma}{|\alpha|}$.

The first three cases cannot be part of the optimal contract: (i) $\lambda_i = 0$, for which the first-order conditions solve for $\Delta w = \frac{Q_H}{1-\alpha}$, and $B_i = B_j = -\frac{(1+\alpha)Q_H}{1-\alpha} < 0$, which contradicts $B_i \geq 0$. (ii) $\Delta w = B_i = B_j = 0$, contradicts $\lambda_1 \geq 0$ and $\frac{d\pi}{d\Delta w} \leq 0$ being valid at the same time. (iii) $\frac{d\pi}{d\Delta w} = 0$, $\lambda_1 = 0$, and $B_i = B_j = 0$ give $\Delta w = \frac{1}{4}Q_H$, but also gives: $\frac{d\pi}{dB_i} = \frac{d\pi}{dB_j} = \frac{1}{4}(1-\alpha)Q_H + \lambda_{2,4}$, which contradicts $\frac{d\pi}{dB_i} = \frac{d\pi}{dB_j} \leq 0$, as $-1 < \alpha < 1$. The next three cases jointly characterize the optimal contract: (iv) $\frac{d\pi}{dB_i} = \frac{d\pi}{dB_j} = 0$ and $\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0$ give $B_j = \frac{1}{2}Q_H$, all conditions are satisfied as long as $\frac{\gamma}{|\alpha|} \geq \frac{1}{2}Q_H \geq 0$, otherwise B_i will hit either constraint, which leads to contradiction. This case leads to the first line

in Proposition 1. (v) $\frac{d\pi}{d\Delta w} = 0$, and $\lambda_1 = 0$ gives $\Delta w = \frac{1}{4}Q_H - \frac{(3+\alpha)\gamma}{4(1+\alpha)|a|}$, all conditions are satisfied as long as $\gamma < \frac{(1+\alpha)|\alpha| Q_H}{3+\alpha}$ otherwise $\Delta w \leq 0$ which gives a contradiction. This case gives the third line in Proposition 1. (vi) $\frac{d\pi}{d\Delta w} \leq 0$, and $\lambda_1 \geq 0$ give $\gamma \geq \frac{(1+\alpha)|\alpha| Q_H}{3+\alpha}$, otherwise the two cannot be satisfied simultaneously, and $\frac{d\pi}{dB_i} = \frac{d\pi}{dB_j} = 0$ together with $\lambda_2 = \lambda_4 = 0$, and $\lambda_3 = \lambda_5 \geq 0$ give that $\frac{\gamma}{|\alpha|} \leq \frac{1}{2}Q_H$ otherwise a contradiction follows. This gives the second line in Proposition 1 and completes the characterization of the optimal contract.

To complete the proof of the optimal contract we need to compare the profits under truthful peer evaluation with the profits of peer evaluation that is not truthful. For non-truthful peer evaluation under good co-worker relations ($m_{ij} = 1$), the principal always has to pay a higher peer evaluation bonus, while incentives are weaker. Therefore, it is evident that this decreases the principal's profits. The non-truthful evaluation strategy for bad co-worker relations ($m_{ij} = 0$) seems more promising, as the principal never has to pay the peer evaluation bonus. The size of the peer evaluation bonus therefore does not matter for profits, it should only be sufficiently high so as to give rise to the non-truthful peer evaluation strategy. The principal still needs to find the optimal team bonus. For $m_{ij} = 0$ the agent's optimal effort is:

$$e_i = (1+\alpha)\Delta w - \alpha\gamma,$$

leading to:

$$\frac{d\pi}{d\Delta w} = (1+\alpha)\left[Q_H - 4\Delta w\right] + 2\alpha\gamma = 0,$$

an optimal team bonus under $m_{ij} = 0$:

$$\Delta w = \tfrac{1}{4}Q_H + \frac{\alpha\gamma}{2(1+\alpha)} \text{ if } \alpha\gamma > -\frac{1}{2}Q_H(1+\alpha),$$
$$\Delta w = 0 \text{ if } \alpha\gamma < -\frac{1}{2}Q_H(1+\alpha),$$

and giving profits:

$$\pi = \tfrac{1}{4}(1+\alpha)Q_H^2 - \alpha\gamma Q_H + \frac{\alpha^2\gamma^2}{(1+\alpha)} \text{ if } \alpha\gamma > -\frac{1}{2}Q_H(1+\alpha),$$
$$\pi = -2\alpha\gamma Q_H \text{ if } \alpha\gamma < -\frac{1}{2}Q_H(1+\alpha).$$

The profits under the weakest form of truthful peer evaluation ($\gamma < \frac{(1+\alpha)\,|\alpha|\,Q_H}{3+\alpha}$) are described by:

$$\pi = \tfrac{1}{4}(1+\alpha)Q_H^2 + \frac{(1-\alpha)\gamma}{2\,|a|}Q_H + \frac{(1-\alpha)^2\gamma^2}{4(1+\alpha)\,|a|^2},$$

which are strictly higher than the profits under $m_{ij} = 0$ with $\alpha\gamma > -\frac{1}{2}Q_H(1+\alpha)$:

$$\pi(m_{ij} = s_{ij}) - \pi(m_{ij} = 0) = \frac{(1-\alpha)\gamma}{2\,|a|}Q_H + \frac{(1-\alpha)^2\gamma^2}{4(1+\alpha)\,|a|^2} + \alpha\gamma Q_H - \frac{\alpha^2\gamma^2}{(1+\alpha)} > 0$$
$$\Rightarrow \left[\frac{1}{2}(1-\alpha)\frac{\gamma}{|\alpha|} + \alpha\gamma\right] Q_H + \frac{1}{(1+\alpha)}\left[\left(\frac{1}{2}(1-\alpha)\frac{\gamma}{|\alpha|}\right)^2 - (\alpha\gamma)^2\right] > 0$$
$$\Rightarrow \frac{1}{2}(1-\alpha)\frac{\gamma}{|\alpha|} > -\alpha\gamma \Rightarrow \gamma\left[1-\alpha+2\alpha\,|\alpha|\right] > 0 \Rightarrow 1-\alpha+2\alpha\,|\alpha| > 0$$
$$\Rightarrow \alpha > -1,$$

which is satisfied by assumption. In case $\alpha\gamma < -\frac{1}{2}Q_H(1+\alpha)$, the profits for non-truthful peer evaluation can be higher than those under truthful peer evaluation. The principal does not have to pay any wages in this case. He exploits the bad relationship between the agents by awarding a high peer evaluation bonus, which he never has to pay. Agents exert effort to increase their disliked colleague's lying costs. For extreme lying costs, the principal can obtain high output with certainty for free.

Comparative statics for expected profits and expected utility

The expected profits under truthful peer evaluation can be written as:

$$E(\pi) = 2\left[(1+\alpha)\Delta w + B\right]\left[Q_H - 2\Delta w - B\right]. \tag{B1}$$

By substituting the optimal contract into (B1) gives:

$$E(\pi) = \frac{1}{2}Q_H^2,$$

for contract (A), leading to the comparative static effects :

$$\frac{\partial E(\pi)}{\partial \gamma} = \frac{\partial E(\pi)}{\partial \alpha} = 0.$$

Similar, for contract (B) we obtain:

$$E(\pi) = 2\frac{\gamma}{|\alpha|}\left[Q_H - \frac{\gamma}{|\alpha|}\right],$$

and:

$$\begin{aligned}\frac{\partial E(\pi)}{\partial \gamma} &= \frac{2}{|\alpha|}\left[Q_H - 2\frac{\gamma}{|\alpha|}\right] > 0,\\ \frac{\partial E(\pi)}{\partial \alpha} &= -\frac{2\alpha\gamma}{|\alpha|^3}\left[Q_H - 2\frac{\gamma}{|\alpha|}\right],\end{aligned}$$

as comparative static results. Expected profits are increasing in γ for contract (B), as $\frac{\gamma}{|\alpha|} < \frac{1}{2}Q_H$. Likewise, expected profits are increasing in α for $\alpha < 0$, and decreasing in α for $\alpha > 0$, i.e., expected profits decrease with more pronounced co-worker relation. The expected profits under contract (C) are:

$$E(\pi) = 2\left[\frac{(1+\alpha)Q_H}{4} + \frac{(1-\alpha)\gamma}{4\,|a|}\right]\left[\frac{Q_H}{2} + \frac{(1-\alpha)\gamma}{2(1+\alpha)\,|a|}\right],$$

which gives comparative static results:

$$\begin{aligned}\frac{\partial E(\pi)}{\partial \gamma} &= \frac{(1-\alpha)Q_H}{2\,|a|} + \frac{(1-\alpha)^2\gamma}{2(1+\alpha)\,|a|^2} > 0,\\ \frac{\partial E(\pi)}{\partial \alpha} &= -\frac{\alpha\gamma}{|\alpha|^3}\left[\frac{(1-\alpha)Q_H}{2} + \frac{(1-\alpha)^2\gamma}{2(1+\alpha)\,|\alpha|}\right] + 2\Delta w\left[Q_H - 2\Delta w - B\right].\end{aligned}$$

Expected profits are increasing in γ, as lying costs shift compensation towards a higher peer evaluation bonus, which is a more cost-effective manner of stimulating effort. Co-worker relations have two effects on profits: First, as before, co-worker

relations affect expected profits through a change in the peer evaluation bonus. Second, better co-worker relations increase the effectiveness of the team bonus. Expected profits are increasing in α, except for moderately positive co-worker relations, as the effect through the peer evaluation bonus dominates the favorable effect on the free-rider problem for a small team bonus.

The comparative static effects with respect to expected utility are best shown when we rewrite (3.3) slightly. A substitution of U_j into i's utility function, in combination with the homogeneity assumption, gives:

$$\begin{aligned} E(U) &= \frac{1}{1-\alpha}\left[\varphi(\cdot)\Delta w + p(\cdot)B - C(\cdot)\right] \\ &= \frac{1}{1-\alpha}\left[\frac{1}{2}B^2 + 2B\Delta w + \frac{1}{2}(3-\alpha)(1+\alpha)\Delta w^2\right]. \end{aligned}$$

The expected utility under contract (A) is:

$$E(U) = \frac{1}{1-\alpha}\frac{Q_H^2}{8},$$

which yields the following comparative statics:

$$\begin{aligned} \frac{\partial E(U)}{\partial \gamma} &= 0, \\ \frac{\partial E(U)}{\partial \alpha} &= \frac{1}{(1-\alpha)^2}\frac{Q_H^2}{8} > 0. \end{aligned}$$

Contract (B) gives expected utility:

$$E(U) = \frac{1}{1-\alpha}\frac{1}{2}\frac{\gamma^2}{|\alpha|^2},$$

and comparative statics:

$$\begin{aligned} \frac{\partial E(U)}{\partial \gamma} &= \frac{\gamma}{(1-\alpha)\,|\alpha|^2} > 0, \\ \frac{\partial E(U)}{\partial \alpha} &= \frac{\gamma^2}{(1-\alpha)\,|\alpha|^2}\left[\frac{1}{2(1-\alpha)} - \frac{\alpha}{|\alpha|^2}\right]. \end{aligned}$$

In contract (B), the expected utility is increasing in α, except for $0 < \alpha < \frac{2}{3}$. Finally,

the expected utility in contract (C) equals:

$$\begin{aligned} E(U) &= \frac{1}{1-\alpha}\left[\frac{1}{2}B^2 + 2B\Delta w + \frac{1}{2}(3-\alpha)(1+\alpha)\Delta w^2\right] \\ &= \frac{1}{32(1-\alpha)}\left[(3-\alpha)Q_H - \frac{(5+\alpha)(1-\alpha)\gamma}{(1+\alpha)\,|\alpha|}\right]\left[(1+\alpha)Q_H + \frac{(1-\alpha)\gamma}{|\alpha|}\right] \end{aligned}$$

and comparative static analysis gives:

$$\begin{aligned} \frac{\partial E(U)}{\partial \gamma} &= -\frac{1}{1-\alpha}\frac{\partial B}{\partial \gamma}\left[\frac{(1-\alpha)B}{2(1+\alpha)} + \frac{(1-\alpha^2)\Delta w}{4}\right] < 0 \\ \frac{\partial E(U)}{\partial \alpha} &= -\frac{1}{1-\alpha}\frac{\partial B}{\partial \alpha}\left[\frac{(1-\alpha)B}{2(1+\alpha)} + \frac{(1-\alpha^2)\Delta w}{4}\right] + \frac{B\left[(3-\alpha)(1+\alpha)\Delta w + 2B\right]}{2(1-\alpha)^2} \\ &\quad + \frac{1}{(1-\alpha)^2}\left[\frac{1}{2}B^2 + 2B\Delta w + \frac{1}{2}(3-\alpha)(1+\alpha)\Delta w^2\right] \end{aligned}$$

Expected utility is decreasing in the lying costs. Although higher lying costs allow for a higher peer evaluation bonus, which increases rents as we saw in contract (B), the team bonus decreases simultaneously, which has a stronger effect on worker's rents. There is a similar effect on utility for co-worker relations. As these become more pronounced, compensation shifts towards a higher team bonus, denoted by the first term. This term is negative for $\alpha < 0$ and positive for $\alpha > 0$. Further, utility increases in the co-worker relations through a softened free-rider problem and the direct effect on utility, captured by the second and third term respectively. In sum, utility is increasing in α, excepts for some moderately negative co-worker relations, where the first term dominates.

Chapter 4

A Field Experiment on Team Incentives and Peer Pressure

Joint with Josse Delfgaauw, Robert Dur, Okemena Onemu and Willem Verbeke.

4.1 Introduction

Team incentives are a widely used means of stimulating and rewarding employees' performance (Ledfort et al. 1995, CIPD 2010). Sometimes, the nature of the production process makes that team-based pay is the best or only way to reward employees for performance, for instance when team performance is much easier to assess than individual performance. In other cases, organizations use team incentives so as to avoid inequity and rivalry at the workplace that individual performance pay may bring about.

A well-known problem with team incentives is free-riding (Holmstrom 1982). That is, each individual employee has a too strong incentive to slack off as the benefits from working hard (the team rewards) are shared with all team members, while the costs of working hard are borne individually.[1]

[1]Perhaps the cleanest evidence of such free riding in a field context is provided by Erev et al. (1993) who study incentives for orange-pickers in Israel. They show that production is 25% higher under individual performance pay than under an equivalent incentive based on performance of a team of four workers. See also Gneezy et al. (2004).

The literature has suggested several ways in which employees can alleviate such free-rider problems, ranging from punishing or pressuring shirking colleagues to encouraging and helping them. Employees who observe the behavior of their peers can exercise pressure on those colleagues that do not reach the group norm, e.g., by shaming or other forms of punishment (Kandel and Lazear 1992). Thus, peer pressure can make groups more productive. However, as most people do not enjoy the act of punishing (let alone enjoy being punished), peer pressure also involves a cost to employees (Barron and Gjerde 1997). Employees also have more friendly options to alleviate free-riding in teams. For instance, helping a less productive colleague can foster cooperation (Fitzroy and Kraft 1986, Drago and Turnbull 1988). Besides helping, employees may actively invest in the quality of co-worker relations, as colleagues who care about one another have less incentive to free ride (Rotemberg 1994, Chapter 2). The two extremes of such anti-shirking behavior, peer pressure and relationship-building, can have opposite implications for employee welfare, and for co-worker relations in particular.

This chapter presents a field experiment designed to test a few basic hypotheses on team incentives, peer pressure, and co-worker relations. We ran a field experiment in a retail chain consisting of 128 stores. We introduced short-term team incentives in a random sample of these stores. To assess co-worker relationships and employees' inclinations to exert peer pressure, we conducted questionnaires among employees in all stores both before the experiment started and after the experiment ended. We examine whether the effect of team incentives on performance increases in the quality of co-worker relations prior to the experiment and in employees' pre-experimental stated willingness to take action upon observing a shirking colleague. Further, by comparing questionnaire responses before and after the experiment, we examine how team incentives affect the quality of co-worker relations and employees' intentions to exert peer pressure in the future. Lastly, using these exogenously induced variations, we study how peer pressure and the quality of co-worker relations affect employees' performance.

Our results are as follows. First, we do not find a significant average treatment effect of the team incentive on sales performance. Next, we allow for heterogeneous

treatment effects. Our findings suggest that the team incentive has a stronger effect for stores with a strong anti-shirking culture or better co-worker relations. Even though this observation is in line with the theories set forth above, it cannot be considered as evidence under conventional measures of statistical significance.

Unfortunately we cannot conclude much from our first finding that the team incentive had no significant effect on sales performance, because, with hindsight, the targets we set were most likely to be too ambitious. Stores in the treatment could earn a team bonus by increasing the number of goods sold per purchase above a store specific target. This target was set 10% above the average number of goods sold per purchase during the previous month. In retrospect this target may have been too ambitious. For stores in the control group, the number of goods sold per purchase were lower as compared to the month used to set the targets. These unfavorable circumstances added to the difficulty of reaching the target. Only one out of the sixty stores in the treatment managed to obtain the team bonus, making it questionable whether the team bonus gave stores an incentive to increase sales effort. Our additional findings also have to be seen in this light.

Our second finding comes from comparing questionnaire responses before and after the experiment. We find that the quality of co-worker relations significantly deteriorated in response to the treatment. Further, the short-run team incentive led to a significant reduction in post-experimental willingness to exert peer pressure. Our questionnaire data also hint at the reason for this reduction. Employees in the team bonus treatment report more often that the colleague to whom peer pressure was directed reacted with remorse and that it less often led to an improvement of the colleague's behavior. Taken together, it seems plausible that employees in treated stores exercised more peer pressure during the experiment with unsatisfactory results, lowering the willingness to undertake such behavior in the future. It is unlikely that we picked up a general disappointment for treated stores in the second questionnaire, as we show that job satisfaction and employee-manager relations were not affected by the temporary team incentive.

Third, our design arguably allows for an investigation of the effect of peer pressure and co-worker relations on performance. The treatment can be seen as an

instrument for peer pressure and co-worker relations, as it should be unrelated to post-experimental performance. However, we do not find any difference in post-experimental performance between the control group and treated stores.

There are a number of empirical studies on the effects of peer pressure that are related to our study. In an observational study, Mas and Moretti (2009) make use of detailed scanner data from a supermarket chain and show that social pressure increases productivity. The authors exploit information on the spatial arrangement of cashier seating and show that cashiers raise their productivity when they are being observed by a more productive colleague. The increase cannot stem from pro-social behavior, as the productivity of colleagues a cashier observes turns out not to matter for productivity. Falk and Ichino (2006) find a similar peer effect in a controlled field experiment, where students fill envelopes, either alone or in pairs. They find that productivity is higher in the pairs treatment, and more similar within a pair; that is, the standard deviations of productivity are lower within pairs than within hypothetical pairs formed from subjects in the singles treatment. Bandiera et al. (2010) study workers at a fruit farm and find that workers who are socially connected tend to conform to a common productivity norm. On average, social ties among workers turn out not to affect productivity.

The questionnaire that we use is for a large part based on Freeman et al. (2010). Freeman et al. (2010) included the questions in two surveys: The General Social Survey (2002 and 2006) held among a representative sample of workers throughout the US, and a special NBER survey among 14 companies that have some form of group incentives.[2] In both surveys, a majority of the workers state that they would take action upon observing a shirking colleague. Freeman et al. (2010) find that such behavior is more likely when team incentives are present. Further, they find in establishments where employees are more inclined to take action, employees also report that colleagues work harder. Whether this is a causal relation, however, is not obvious from non-experimental data. Our study extends the analysis by Freeman et al. (2010) by exploiting our experimental setup. Further, we use actual performance data rather than (self-)reported performance data.

[2]Freeman et al. (2008) use a similar questionnaire in a single multinational firm to examine cross-country differences in labor practices.

We proceed as follows. In the next section, we describe the experimental set-up and Section 4.3 provides the data description. We briefly describe the methodology of our empirical analysis in Section 4.4 and the results are reported in Section 4.5. Section 4.6 concludes.

4.2 Experimental set-up

The field experiment took place during the Fall of 2009 within a retail chain that sells clothing, shoes, and sports apparel. The retail chain consists of 128 stores; all stores operate under one brand and are geographically dispersed throughout The Netherlands. The company employed around 1300 store employees during our experimental period. Most of the store employees are sales clerks, whose primary tasks consist of assisting customers to make purchases. Store employees earn a fixed hourly wage slightly above the legal minimum, and are not provided with any monetary incentives other than the rewards provided during the experiment. Store managers earn about 50% more than regular employees, and their pay is partly determined by store performance. Recently, the company's management started to explore possibilities to offer incentives to all employees. The results of an earlier experiment with incentives for store employees in the same retail chain are described in Chapter 5.[3]

In the current experimental treatment, stores could earn a team bonus by raising the average number of items sold per purchase over a six week period. The number of items sold per purchase has a low volatility in comparison to other sales performance measures, which made it an attractive measure to base an absolute target on.[4] Futhermore, this measure gives employees a clear idea how they can contribute. Treated stores could earn one of two monetary rewards by increasing the average number of items sold per purchase by 10 or 20 percent as compared to their previous

[3]The earlier experiment explored the effects of sales contests among stores, and took place in the Fall of 2007 and the Spring of 2008. We don't expect any carry-over effects from the previous experiments, as there is a time-lag of more than a year between both experiments.

[4]The objective of the company's management was to increase profits by stimulating sales performance. However, sales are highly volatile on a week to week basis, which makes it more difficult to determine targets.

month's average. The rewards for surpassing the respective targets were set to 75 euro and 200 euro for each store employee, including the store manager.[5]

We used a stratified randomization procedure to assign stores with a similar number of employees to either treatment or untreated control. We assigned 60 stores to the treatment condition, and the remaining 68 stores served as the control group.[6] Ideally we would stratify with respect to the anti-shirking responses of the first questionnaire. However, this was infeasible, as we desired to measure the anti-shirking behavior as close to the start of the experiment as possible. In fact, the announcement of the experiment took place just before the deadline for the first questionnaire. Instead, we stratified with respect to store size, as size may be related to the intensity of anti-shirking behavior.

The experiment was announced to store managers in a letter signed by the management of the company. Treatment stores received a letter explaining the rules of the short-run team incentive. Stores in the control group only took notice that there was a sales plan in other stores. The letters were sent through the retail chain's regular communication channels to make sure that neither store managers nor store employees could infer that they were taking part in an experiment. Hence, this experiment can be classified as a natural field experiment (Harrison and List 2004).

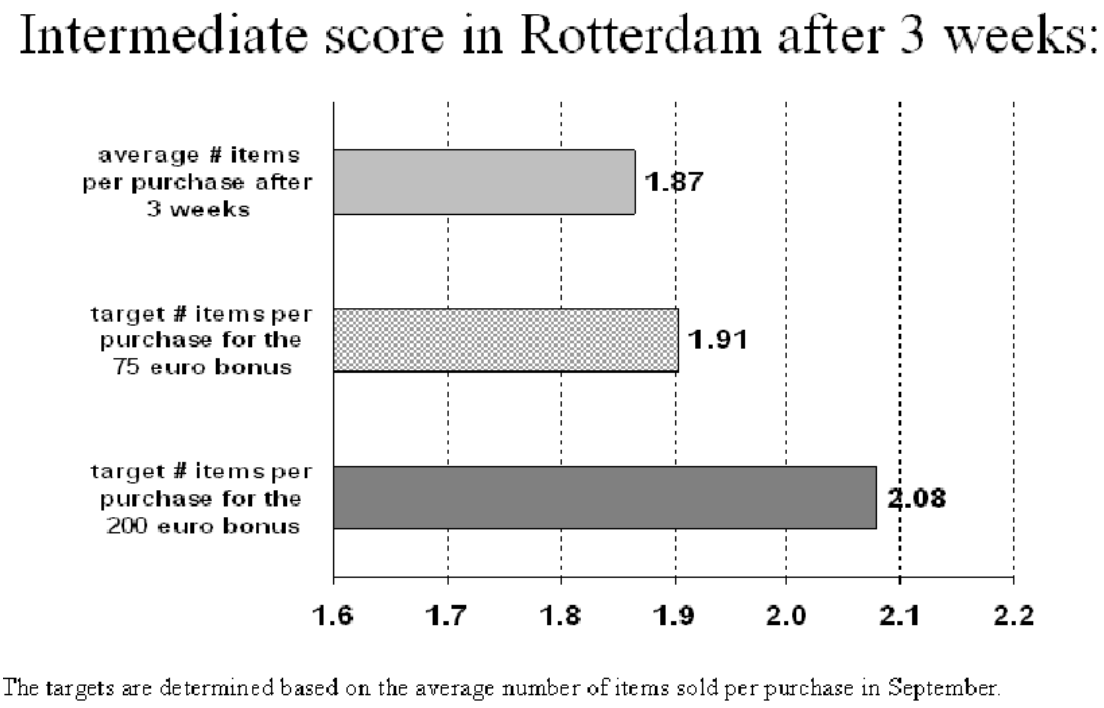
The letter to treated stores also contained a poster. The store manager was instructed to put up this poster in the employee canteen and to replace it with a new copy each week. The posters contained feedback on the average number of items sold per purchase in their store so far and the targets that they had to reach. We used a simple graphical presentation to show how close stores were to reaching their targets, see Figure 4.1 for an example.

All employees within the retail chain were asked to participate in a survey, both prior to the experiment and afterwards. The survey was communicated as an inquiry in employee satisfaction, without mentioning the experiment. For the first survey,

[5] These bonusses would be equivalent to roughly 105 and 280 dollars at the time, and were roughly 5 and 13 percent of an employee's monthly wage. Part-timers would receive half of the reward for both targets.

[6] Stores could be in one of the following five categories: 7 employees or less, 8 and 9, 10 and 11, 12 until 14, or 15 and more employees. This procedure gave 14 (7), 30 (14), 45 (21), 29 (13), and 10 (5) stores in each of the categories respectively, with the number of stores in the treatment between brackets.

Figure 4.1: Example of the feedback poster



prior to the experiment, we sent letters to the private address of store employees asking them to fill out an online survey, which they could enter using a store-identified code. For the second survey, hard-copy questionnaires were sent to the stores. Employees were given the time to complete the questionnaires during their office hours, hoping that this would boost the response.[7]

4.3 Data description

The company provided us with weekly store-level sales data, covering a 64 week period starting from the first week of 2009. It contains indexed turnover figures and the number of items sold per purchase. The experiment took place from week 43 until week 48 in 2009. Additionally, we received a personnel file before and after the

[7]Store managers would receive envelopes containing surveys and pens, and were asked to distribute these among employees. Employees would return their filled out surveys to their store manager in a sealed blank envelope. Managers were instructed to send them back in a marked envelope once a sufficient number of questionnaires had been filled out.

Table 4.1: Descriptive statistics

	All stores		Bonus treatment		Control group		t-test for equality
	Mean	Std	Mean	Std	Mean	Std	of means
Indexed turnover (avarage week 1-42)[a]	98.0	12.1	99.2	11.9	97.0	12.3	-1.02
Number of goods per purchase (wk 1-42)	1.84	0.23	1.86	0.27	1.82	0.19	-1.00
Number of employees	9.6	2.9	9.6	3.0	9.6	2.8	0.16
Percentage of female employees	0.86	0.13	0.84	0.15	0.88	0.12	1.65
Store average age of employees	26.5	4.4	26.4	4.6	26.5	4.4	0.17
Store average tenure of employees	4.7	2.4	4.6	2.3	4.8	2.5	0.45
Gender of the manager	0.42		0.38		0.45		0.73
Age of the managers	39.7	9.5	40.0	9.3	39.4	9.7	-0.35
Tenure of the managers	12.8	8.7	13.3	8.3	12.3	9.0	-0.67
Number of stores	128		60		68		

[a]Turnover index is based on sales in week 43 of 2008.
The first personnel file contained information on 122 store managers, 55 in the bonus, and 67 in the control
t-values are reported, where ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

experiment. The personnel files list for each store the number of employees, their gender, age, tenure, and some information on the type of contract.

Table 4.1 presents the descriptive statistics of sales and the personnel file we received prior to the experiment. In the 42 weeks preceding the experiment an average store sells 1.8 items per purchase and average turnover is close to the base week (week 43 of 2008). Further, a store has 9.6 employees on average, excluding the store manager. Store employees are predominantly female (86%), range in age from 14 to 64, with an average age and tenure of 27 and 5 years, skewed towards younger employees with a median age and tenure of 22 and 2 years respectively. The store manager is slightly more often male (58%) than female, has on average a tenure of 12 years, and is on average 40 years old. The second personnel file is nearly identical and therefore not displayed. Table 4.1 also provides a randomization check: In line with our stratified randomization procedure, the number of employees does not differ between treatment and control, neither do any other observable characteristics.

Figures 4.2 and 4.3 depict the average weekly sales performance of the number of items sold per purchase and the turnover figures respectively, differentiated by treatment and control. The number of goods sold per purchase has a lower volatility on a week to week basis compared to the turnover figures, which was why we chose this measure to determine targets during the experiment. Figure 4.2 shows that

treated stores sell slightly more items per customer over most of the sample period, despite the random assignment of stores. This underlines the importance of including store fixed effects in our estimations of the treatment effects.

Figure 4.2: Average numer of goods sold per purchase

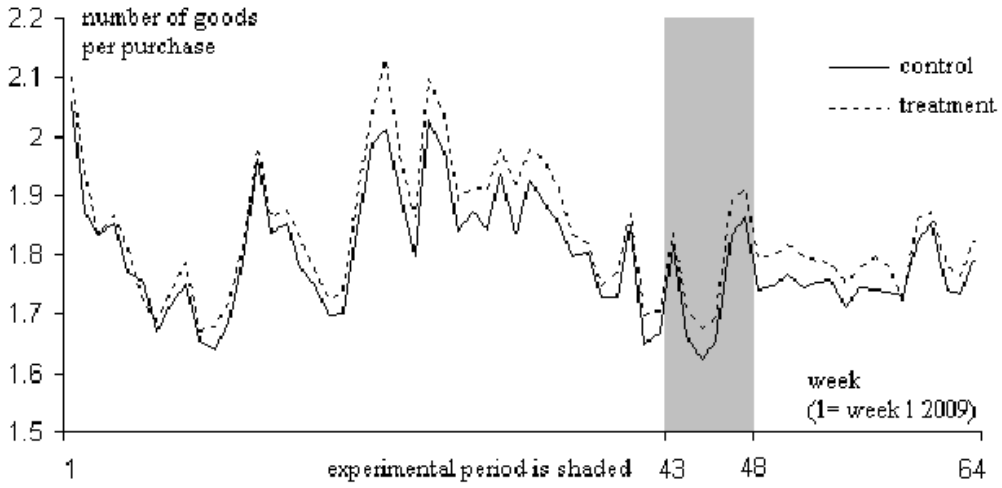
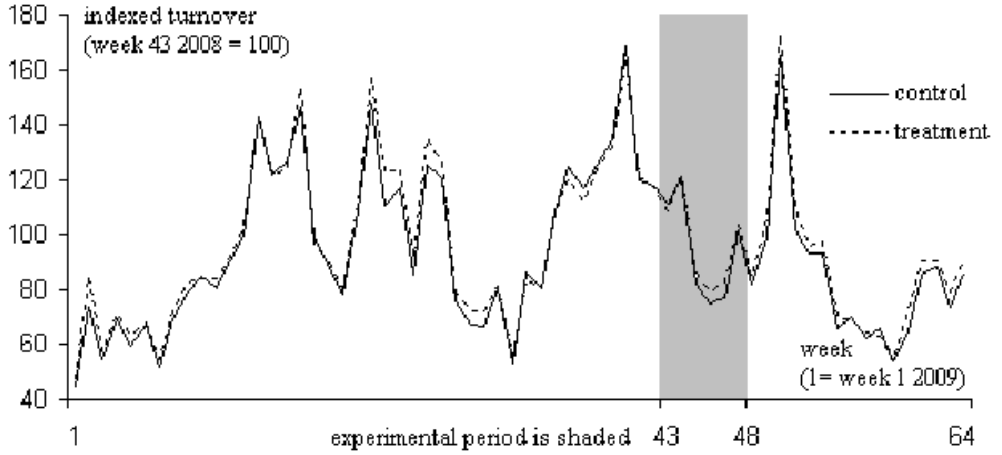


Figure 4.3: Average turnover figures



Apart from sales performance, we also make use of 'soft' variables that are constructed from the data gathered with questionnaires. The first survey, where employees were approached at their private addresses, had a response rate of 31%. For the second survey we gave employees the opportunity to fill out the questionnaire at work, which led to a response rate of nearly 43%. However, the first survey method yielded a response from 117 stores, whereas in the second survey only 77 stores returned the questionnaires. The response rates and some characteristics of the respondents can be found in Table 4.2 together with the descriptive statistics of both questionnaires. It is possible that the different methods attracted a different sample of employees, e.g., respondents in the first survey are likely to be more committed to the company, as they completed the survey in their own time. This may give rise to a bias in our results, but only when sample selection is correlated with the treatment. We found that the hours of work differs significantly between the two questionnaires, where repondents in the second questionnaire work more hours. However, this difference is not correlated with the treatment. Table 4.2 also shows that respondents in the first questionnaire were significantly younger in the treatment than in control stores. This difference is not present in the second questionnaire. We include individual controls in the difference-in-difference analysis of the questionnaire data to prevent that age differences in response bias the estimate of the treatment effect.

The questionnaires included statements on job satisfaction, co-worker relations, the relationship with the store manager, and anti-shirking behavior. Employees could indicate to what extent they agreed to statements on a 7-point Likert scale, ranging from "strongly disagree" to "strongly agree". The wording of the statements and their combination into variables of interest can be found in the Appendix. Table 4.2 shows for all variables their mean, standard deviation, and t-test for the equality of means between treatment and control within a survey. There are some noteworthy differences in the questionnaire prior to the experiment: Stores in the treatment happen to be significantly more positive about their co-worker relations, more likely to talk to a shirking colleague, and more positive about the consequences of anti-shirking behavior. These differences disappeared in the second questionnaire.

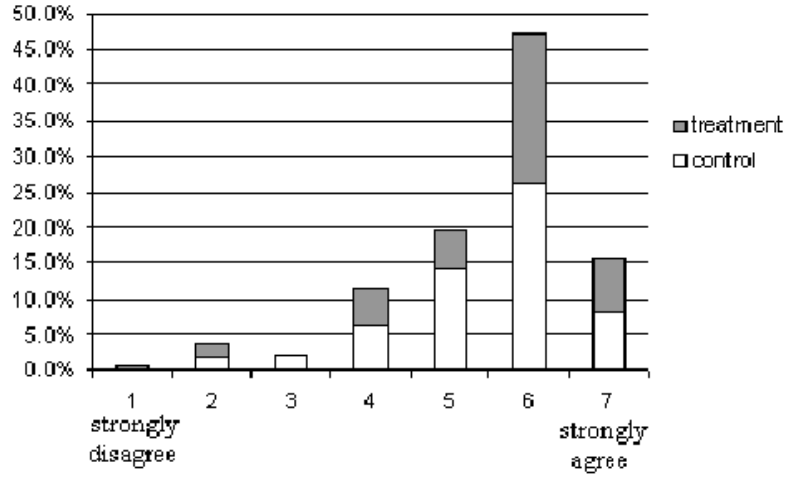
Table 4.2: Descriptive statistics of both questionnaires

	questionnaire 1							questionnaire 2						
	All stores		Bonus treatment		Control group		t-test for equality	All stores		Bonus treatment		Control group		t-test for equality
	Mean	Std	Mean	Std	Mean	Std	of means	Mean	Std	Mean	Std	Mean	Std	of means
Respondents (#)	373		156		217			506		216		290		
Response (%)	31		28		34			43		39		46		
Stores (#)	117		52		65			77		35		42		
Stores with response >20% (#)	90		36		54			77		35		42		
Female (%)	88		89		86		0,79	84		84		84		0,02
Age category (%): <20 years	37		41		33		-0.85	37		36		37		-0.14
20 - 30 years	35		40		30		-1.86 *	38		38		37		0.08
>30 years	29		19		36		3.01 ***	26		26		26		0.08
Tenure category (%): <1 year	24		27		22		-0.33	24		27		23		-1.90 *
1 - 5 years	35		37		34		-0.72	36		31		40		2.24 **
5 -10 years	25		20		28		1.30	23		25		21		-0.24
>10 years	16		16		16		-0.11	17		17		16		0.17
Hours contracted category (%): <10 hours a	20		22		18		-0.34	16		16		17		0.08
10 - 20 hours a week	31		28		34		0.94	29		31		26		1.09
20 - 30 hours a week	26		25		26		-0.06	24		23		25		-0.54
> 30 hours a week	23		24		22		-0.58	32		31		33		-0.64
Co-worker relations	17.26	2.95	17.71	2.75	16.92	3.06	-2.35 **	17.66	2.24	17.55	2.19	17.73	2.28	0.91
Job satisfaction	16.9	3.55	17.11	3.55	16.75	3.55	-0.89	17.28	3.25	17.59	2.94	17.06	3.45	-1.79 *
Relationship with manager	4.92	1.78	5.05	1.79	4.83	1.77	-1.16	5.42	1.44	5.42	1.47	5.43	1.42	0.04
Observability colleague's effort	5.5	1.2	5.56	1.29	5.46	1.14	-0.71							
Anti-shirking index:	17.41	4.07	17.89	4.2	17.06	3.95	-1.84 *	17.32	4.47	17.09	4.64	17.48	4.35	0.96
- Do nothing	3.87	1.81	3.79	1.8	3.93	1.82	0.71	3.95	1.93	4.04	1.97	3.89	1.9	-0.85
- Talk to the colleague in private	3.86	1.86	4.13	1.87	3.67	1.82	-2.26 **	4.07	1.96	4.1	2,00	4.05	1.94	-0.33
- Talk to the supervisor	4.9	1.69	4.91	1.71	4.89	1.68	-0.10	4.95	1.74	4.92	1.78	4.98	1.72	0.40
- Talk to other colleague	4.54	1.65	4.64	1.63	4.47	1.67	-0.93	4.26	1.71	4.13	1.67	4.35	1.74	1.44
Consequences anti-shirking:	19.08	3.84	19.73	3.16	18.62	4.2	-2.28 **	18.76	3.64	18.89	3.76	18.66	3.55	-0.66
- Colleague took offence	3.62	1.53	3.4	1.45	3.77	1.56	1.92 *	3.7	1.62	3.72	1.62	3.69	1.63	-0.16
- Other colleagues appreciated it	5.00	1.2	5.11	1.06	4.93	1.29	-1.22	4.82	1.36	4.87	1.37	4.78	1.34	-0.75
- Store manager appreciated it	4.92	1.41	5.02	1.39	4.85	1.42	-0.93	4.94	1.36	5,00	1.41	4.89	1.32	-0.81
- Colleague improved behavior	4.78	1.29	5.02	1.19	4.62	1.34	-2.50 **	4.73	1.41	4.78	1.49	4.69	1.35	-0.67

t-values are reported, where ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

A prerequisite to act against shirking colleagues is the ability to observe their behavior. In the first questionnaire we asked employees to indicate their agreement to the statement: "It is easy to see whether co-workers are working well or poorly". Figure 4.4 shows a histogram of their responses. Most employees state that they either agree or strongly agree (63% combined) with this statement.

Figure 4.4: Distribution of the responses to "It is easy to see whether co-workers are working well or poorly".



Given that most employees are able to observe whether a colleague is working hard or not, they can act against a shirking colleague. We looked into employees' willingness to undertake four possible actions upon observing a colleague that slacks off. We asked how likely they would be to "do nothing", "talk to the colleague in private", "speak with the store manager", and "talk about it with other colleagues". The anti-shirking index is formed by adding the scores of the four statements, where the score from "do nothing" is reversed.[8] Figures 4.5 and 4.6 show a histrogram of the anti-shirking index for the first and the second questionnaire, differentiated by control and treatment. Figure 4.5 illustrates that the distribution of the anti-shirking index is relatively more concentrated towards the higher end for treated employees, whereas Figure 4.6 shows no such difference for the questionnaire after the experiment.

[8]We find a low Cronbach alpha coefficient, which suggest that this index is a construct of different variables. However, we chose to conform to the anti-shirking index in Freeman et al. (2010), such that we can compare our findings to theirs.

Figure 4.5: Distribution of the anti-shirking index in the first questionnaire

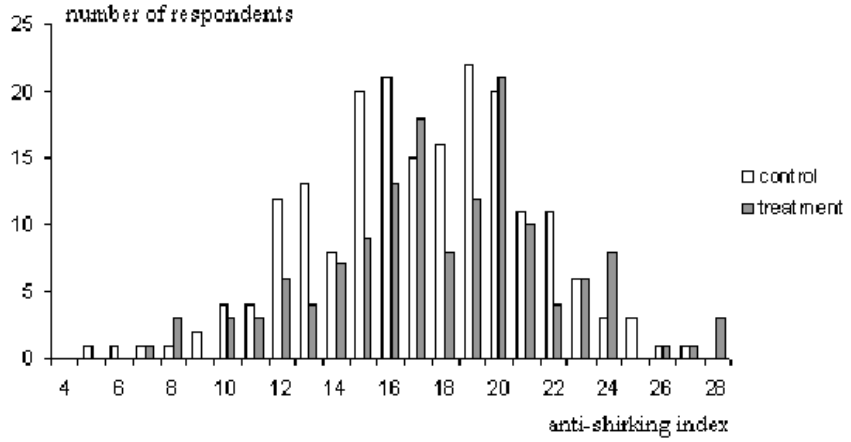
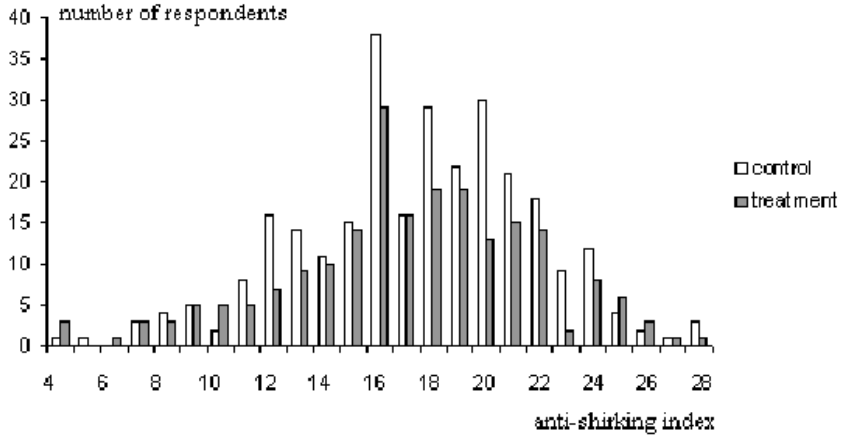


Figure 4.6: Distribution of the anti-shirking index in the second questionnaire



4.4 Method

We use OLS panel estimation including week and store fixed effects to estimate the treatment effect. We estimate the treatment effect on the number of items sold per purchase, as the bonus made this performance measure most salient during experiment. The regression equation reads:

$$y_{st} = \alpha_s + \theta_t + \gamma T_{st} + \varepsilon_{st} \tag{4.1}$$

Let y_{st} be the logarithm of the items sold per purchase, and T_{st} be a dummy variable which is equal to one during the experimental period t when store s is assigned to the bonus treatment. Store fixed effects are indicated by α_s, week fixed effects by θ_t, and ε_{st} is an error term.[9] The treatment effect is given by γ. Next, we allow the treatment effect to be heterogeneous in the extent to which stores may suffer from free-riding. This gives the following regression:

$$y_{st} = \alpha_s + \theta_t + \gamma T_{st} + \delta T_{st} X_s + \mu P_t X_s + \varepsilon_{st} \tag{4.2}$$

where X_s can be the anti-shirking index, co-worker relations, and the number of employees of store s, measured before the experiment. Let P_t be a dummy variable that has a value of one during the experimental period. By interacting P_t with X_s we make sure that period-specific effects of X are not erroneously picked up by the interaction with the treatment (δ). We also present the results when we take the indexed turnover figures as the dependent variable. However, this robustness check informs us perhaps more about the production function than about the treatment effect.

We estimate the treatment effect on the variables from the questionnaires with a difference-in-difference approach. A change in responses between the questionnaires before and after treatment may have occured in both the treatment and the control group. We identify the treatment effect as the difference between those changes, and estimate this using the following equation:

$$q_{ist} = \mu + \beta T_s + \gamma S_t + \lambda T_s S_t + \varepsilon_{ist} \tag{4.3}$$

Let q_{ist} be the response to a particular item by employee i from store s in questionnaire t, where q can be co-worker relations, employee-manager relations, job satisfaction, or anti-shirking behavior. The items can take a value of 1 to 7 on a Likert-scale indicating to what extent an employee agrees or disagrees with a statement in the questionnaire, or in some cases q corresponds to the value of a summation of responses to a number of statements. The combination of items into variables can

[9] The standard errors are robust to serial correlation, as specified by Arellano (1987).

be found in the Appendix. Let T_s be a dummy variable for individuals in the treated stores, and S_t be a dummy variable for responses in the second questionnaire. The error term is indicated by ε_{ist}.[10] The parameter λ gives the difference-in-difference estimator, i.e., the treatment effect on q. We report the results of the estimations using OLS, as they are easier to interpret than ordered choice probit.[11]

In order to estimate whether items from our questionnaire have an effect on sales performance, we shift our attention to the turnover during 16 weeks after the experiment. Outside of the experimental period we focus on turnover figures, as they give a more complete picture of store performance than the number of items sold per purchase. This is also reflected in the store managers' compensation, which depends partially on the realized turnover. First, we estimate:

$$y_{st} = \alpha_s + \theta_t + \gamma T_s N_t + \varepsilon_{st} \tag{4.4}$$

where N_t is a dummy variable that takes the value of one in the 16 weeks after the experiment, and T_s is a dummy variable that takes a value of one for stores that were treated. In case we observe an effect of the treatment after the experiment, we may be able to attribute this effect to a change in one of the items in the questionnaire. Therefore, we include the changes in q to see whether this explains the lagged effect of being treated:

$$y_{st} = \alpha_s + \theta_t + \gamma T_s N_t + \delta \Delta q_s N_t + \varepsilon_{st}$$

where Δq_s can for example indicate the change in anti-shirking behavior between the first and second survey for store s. The effect of anti-shirking behavior on performance could suffer from an endogeneity problem in a cross-sectional analysis, as was argued in the introduction. We exploit our experimental treatment to avoid this problem. We measured anti-shirking behavior both before and after the experiment, where it may been affected by the treatment. Under the assumption that effect of

[10] The analysis of the questionnaire data can be described as a cluster-randomized experiment, because we draw multiple observations from most stores. Therefore, we allow the standard errors to be clustered on the stores level. Including store random effects does not affect the results qualitatively.

[11] We perform ordered probit estimation as a robustness check. The results do not depend on the estimation procedure, as ordered probit yields qualitatively similar results. The ordered probit estimations are available upon request.

anti-shirking on performance is identical before and after the experiment, we can use Δq_s to find the effect of anti-shirking behavior on performance. Namely, the effect of the initial level of anti-shirking behavior on performance, i.e., before the experiment, will be picked up by the store fixed effects α_s. In case the treatment induced exogenous variation in anti-shirking, the coefficient δ captures the effect of anti-shirking behavior on performance.

4.5 Results

We begin by estimating the treatment effect on the number of items sold per purchase. Column 1 of Table 4.3 gives the results of estimating (4.1).[12] We find an insignificant average treatment effect. In other words, the short-run team incentive did not affect the number of items sold per purchase on average. Next, we allow the treatment effects to be heterogeneous to the extent that stores may suffer from free-riding. Free-riding is hypothesized to be less severe for stores with a smaller number of employees, a stronger anti-shirking culture, or better co-worker relations. The results of estimation (4.2), where we include the interaction of the treatment dummy with the variables mentioned, can be found in column 2.[13] None of the interaction terms are statistical significant. Neither do we find a treatment effect on turnover.

In Table 4.4 we turn to the difference-in-difference analysis of the questionnaire data. In our baseline regression, we regress the anti-shirking index on the treatment dummy, second questionnaire dummy, and their interaction. The treatment effect on anti-shirking behavior is the difference between the treatment and control in the difference between the questionnaire before the experiment and the questionnaire afterwards. This diff-in-diff coefficient is given by the interaction of the treatment dummy with the second questionnaire dummy. The results of (4.3) are presented in column 1 of Table 4.4. The treatment had a significant and negative effect on the willingness to undertake anti-shirking behavior. Our estimate is a drop of 1.25

[12] Throughout the thesis, p-values are based on two-sided tests.

[13] The number of employees, anti-shirking index, and co-worker relations reported in the estimations in column 2 and 4 of Table 3 are mean centered.

Table 4.3: The effect of short-run team incentives on sales performance

dependent variable:	ln(number of items sold per purchase)				turnover index			
independent variables	(1)		(2)		(3)		(4)	
treatment	0.004	(0.005)	0.005	(0.005)	-0.300	(1.825)	-1.081	(1.704)
number of employees x treatment			0.002	(0.002)			-0.513	(0.644)
number of employees x experimental period			-0.003	(0.001)**			1.257	(0.508)**
anti-shirking index x treatment			0.003	(0.002)			0.100	(0.753)
anti-shirking index x experimental period			-0.002	(0.001)			0.598	(0.590)
co-worker relations x treatment			0.002	(0.002)			0.207	(0.543)
co-worker relations x experimental period			-0.001	(0.001)			0.082	(0.480)
Period fixed effects	YES		YES		YES		YES	
Store fixed effects	YES		YES		YES		YES	
Number of stores included	128		113		128		113	
Observations	8128		7168		8192		7232	
R^2	0.871		0.879		0.731		0.734	
Log likelihood	12712.9		11300.4		-35710.2		-31411.6	

***, **, * indicate sta tistical significance at the 1%, 5%, and 10% level, respectively.
Number of employees, anti-shirking index, and co-worker relations are mean-centered variables.
Robust standard errors in parentheses

Table 4.4: The effect of short-run team incentives on the anti-shirking index

dependent variable: anti-shirking index						
independent variables	(1)		(2)		(3)	
treatment	0.826	(0.465)*	0.769	(0.435)*	0.782	(0.424)*
questionnaire 2	0.429	(0.347)	0.189	(0.361)	0.215	(0.387)
treatment x questionnaire 2	-1.253	(0.573)**	-1.359	(0.570)**	-1.399	(0.602)**
number of employees			-0.034	(0.064)	0.026	(0.067)
employee-manager relation			0.594	(0.187)***	0.547	(0.145)***
observability effort			0.046	(0.246)	0.157	(0.244)
female respondent					-1.149	(0.359)***
age category (<20 yr excl.)						
20 yr - 30 yr					1.182	(0.478)**
>30 yr					1.817	(0.618)***
tenure category (<1yr excl.)						
1 yr - 5 yr					-0.432	(0.394)
5 yr - 10 yr					0.137	(0.599)
> 10 yr					0.313	(0.627)
hours contracted category (<10 hr excl.)						
10 hr - 20 hr					0.268	(0.470)
20 hr - 30 hr					0.319	(0.540)
>30 hr					0.983	(0.502)**
Number of stores included	122		115		114	
Observations	832		794		739	
R^2	0.005		0.021		0.091	

***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.
Robust standard errors in parentheses

points on the anti-shirking index, where the standard deviation of the anti-shirking index across stores is about 2.2 points.

The reduction in the willingness to undertake anti-shirking for treated stores is robust to the inclusion of a number of store and individual control variables. In column 2 we add the number of employees, the store's average score of the employee-manager relations, and the average score on the ability to observe colleagues (which was present during the first questionnaire), as store control variables. The diff-in-diff estimate remains comparable in size and significance. Store size is not significantly related to anti-shirking behavior, however the biggest store in our sample employs 20 people, while in Freeman et al. (2010) notable difference in anti-shirking arise for firms with more than a hundred employees. In line with Freeman et al. (2010) we find a positive relation between the employee-manager relations and anti-shirking

behavior. Finally, in column 3 the respondent's gender, age category, and hours contracted are added as individual control variables, which does not alter the results either. Male employees, older employees, and full-time employees report a higher willingness to undertake anti-shirking behavior.

We perform the same estimation procedure to find the treatment effect on the respondents' experienced consequences of anti-shirking. The results are presented in column 1 of Table 4.5. Respondents in treated stores are less pleased with the consequences of their anti-shirking behavior after the experiment. The estimated difference is almost 1.0 point on the consequence variable and is significant at the 10 percent level. In columns 2 to 5, we split up the experienced consequences variable into the separate statements. The effect stems mainly from an expressed dissatisfaction on two statements: "The employee not working well resented it" and "The employee not working well improved" with p-values of 0.11 and 0.07 for the diff-in-diff estimates.

In Section 4.1 we argued that anti-shirking behavior can range from pressuring peers to strengthening social ties with colleagues. The extremes of anti-shirking behavior have opposite predictions on how a team incentive affects the quality of co-worker relations. Hence, the effect of the temporary team incentive on co-worker relations allows us to distinguish between the types of anti-shirking behavior. Table 4.6 presents the results of this difference-in-difference estimation in column 1, and in columns 2 and 3 with added controls. Column 1 shows that co-worker relations suffered from the temporary team incentive, as the diff-in-diff estimate is negative and significant at the 5% level, and significant at the 10% level when including all controls. The drop in the quality of co-worker relations in treated stores suggests that peer pressure was the more prevalent type of anti-shirking behavior during the experiment.

In sum, employees in treated stores experienced a drop in the quality of co-worker relations, were less satisfied with the consequences of their anti-shirking behavior after the experiment, and became less willing to exert such anti-shirking behavior in the future. The combination of these results leads us to the conjecture that a dissatisfaction with increased peer pressure in response to the team bonus offers the

Table 4.5: The effect of short-run team incentives on the stated experiences with anti-shirking behavior

	dependent variable: consequences anti-shirking					separate statements				
independent variables	(1)		(2)		(3)		(4)		(5)	
treatment	0.851	(0.450)*	-0.272	(0.184)	0.141	(0.148)	0.079	(0.156)	0.375	(0.181)**
questionnaire 2	-0.286	(0.384)	-0.022	(0.157)	-0.184	(0.133)	-0.081	(0.148)	0.005	(0.162)
treatment x questionnaire 2	-0.955	(0.540)*	0.366	(0.230)	-0.131	(0.196)	-0.071	(0.215)	-0.420	(0.228)*
Store control variables	YES		YES		YES		YES		YES	
Individual control variables	YES		YES		YES		YES		YES	
Number of stores included	111		111		111		111		111	
Observations	654		667		665		666		666	
R^2	0.080		0.051		0.040		0.071		0.049	

***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.
Robust standard errors in parentheses
(1) - Composite of consequence statements (2) - (5)
(2) - The employee not working well resented it
(3) - Other employees appreciated it
(4) - Supervisor appreciated it
(5) - Employee not working well improved

most likely explanation for the differences in the questionnaire responses.

An alternative explanation for our findings is that employees in treated stores expressed a general disappointment in the second questionnaire, as only 1 out of the 60 stores managed to earn the team bonus. We perform another difference-in-difference analysis for both job satisfaction and employee-manager relations to examine such a general disappointment for treated stores. Columns 4 and 5 of Table 4.6 show that the diff-in-diff estimates for both the expressed job satisfaction and the employee-manager relations are far from significant. The significance does not change when we drop store or individual control variables. This suggests that a general disappointment for treated stores cannot explain a decrease in the willingness to undertake anti-shirking, experienced consequences of anti-shirking, or co-worker relations. We cannot rule out a second alternative explanation: Employees may have responded differently to anti-shirking behavior in the presence of the short-run team incentive. For example, colleagues may suspect that anti-shirking behavior was motivated by greed in the presence of a team bonus. This alternative could also explain the negative responses regarding the consequences of anti-shirking.

Finally, we investigate whether the variables measured our questionnaire have an effect on sales performance. Table 4.7 presents the result of estimating (4.4). The effect of being treated does not significantly affect the indexed turnover figures for the 16 weeks after the experiment. This makes it difficult to attribute an effect of being treated on sales after experiment to a change in the variables guaged by questionnaires.[14]

[14]For the subsample of stores that completed both surveys, there is a negative effect of being treated on the sales performance after the experiment. This negative effect is explained rather well by the drop in the anti-shirking index for treated stores. The change in the anti-shirking index is significantly related to the sales performance after the experiment. However, as shown in Table 4.7, this result is not robust to the inclusion of all stores.

Table 4.6: The effect of short-run team incentives on co-worker relations, job satisfaction, and employee-manager relations

	dependent variable: co-worker relations						job satisfaction		employee-manager rel	
independent variables	(1)		(2)		(3)		(4)		(5)	
treatment	0.786	(0.384)*	0.651	(0.377)*	0.695	(0.398)*	0.494	(0.449)	0.271	(0.315)
questionnaire 2	0.811	(0.253)***	0.511	(0.243)**	0.575	(0.252)**	0.323	(0.367)	0.577	(0.194)**
treatment x questionnaire 2	-0.970	(0.409)**	-0.766	(0.386)**	-0.761	(0.397)*	0.072	(0.517)	-0.279	(0.273)
Store control variables	NO		YES		YES		YES		YES	
Individual control variables	NO		NO		YES		YES		YES	
Number of stores included	121		114		114		114		114	
Observations	815		776		742		745		742	
R^2	0.016		0.078		0.092		0.059		0.087	

***, **, * indicate sta tistical significance at the 1%, 5%, and 10% level, respectively.
Robust standard errors in parentheses

Table 4.7: The effect of being treated on sales performance after the experiment

	dependent variable: indexed turnover figures
independent variables	(1)
treatment stores x after experiment dummy	1.578 (4.448)
Period fixed effects	YES
Store effects	YES
Number of stores included	128
Observations	8192
R^2	0.731
Log likelihood	-35708.8

***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.
Robust standard errors in parentheses

4.6 Concluding remarks

In this chapter, we described the results of the introduction of a short-run team incentive in a subsample of stores in a single retail chain. The incentive did not lead to additional sales during the experiment for the treated stores. However, the treatment did affect the employees' willingness to undertake anti-shirking behavior and co-worker relations. These variables were gauged by questionnaires, both prior to and after the experiment. Difference in difference analyses revealed that the temporary team incentive had a negative effect on the expressed willingness to undertake anti-shirking behavior. We argue that this decrease results from a dissatisfaction with increased attempts of peer pressure. We find a significant change in how employees in treated stores experience the consequences of their anti-shirking behavior; they expressed that their colleagues had more resentment towards anti-shirking actions and that the colleague in question improved his behavior less often. Moreover, co-worker relations declined in response to the temporary team incentive. We refute an alternative explanation, that employees in treated stores display a general disappointment in the second questionnaire, by showing that job satisfaction and employee-manager relations are unaffected.

A limitation of our study is that we cannot observe employee behavior during the experiment. We have to rely on survey data before and after the experiment

instead. However, some of the concerns about questionnaires, such as a self-serving bias in responses, are of less importance for this study, as we focus on difference-in-differences. Our results would only be invalid in case a bias is correlated with the treatment. Still, not observing the actual behavior is a serious limitation, as we can only speculate about the mechanism behind our findings. We cannot distinguish our interpretation from some alternatives: For example, anti-shirking behavior may have become less appreciated in the presence of the temporary team incentive, as colleagues may suspect that the actions were motivated by greed.

In the introduction we discussed some studies that showed a positive influence of peer pressure on the sales performance. This implies that organisations should strive to form and nurture a strong anti-shirking culture in order to achieve high levels of performance. Management attention may offer an instrument to do so, as employee-manager relationships are positively related to the willingness to undertake anti-shirking behavior, also found by Freeman et al. (2010). Further, in contrast to Freeman et al. (2010), we found that organisations should be cautious with introducing team incentives, as they can harm the anti-shirking culture. Team incentives may either have a negative impact on how colleagues interpret anti-shirking behavior, rendering these actions less effective, or the negative effect is only present when increased attempts to reduce shirking are ineffective. We cannot pin down the mechanism that caused a drop in the willingness to undertake anti-shirking behavior, as we do not observe actual behavior during the experiment. Further research on this topic is needed.

4.A Appendix

Variable construction

Anti-shirking index: The anti-shirking index is constructed from following four statements: "If you were to see a feelow employee not working as hard as he or she should, how likely would you be to: 1. Do nothing, 2. Talk directly to the employee, 3. Talk to the store manager, 4. Talk about it with other colleagues." (1-7 scale, 1 = very unlikely and 7 = very likely for each option in this statement). The anti-shirking index adds the scores, where the score on "Do nothing" is reversed.

Consequences anti-shirking: "What was the outcome when you undertook some action against a colleague that didnt work as hard as he or she should? 1. The colleague resented it, 2. Other colleagues appreciated it, 3. The store manager appreciated it, 4. The colleague not working well improved." (1-7 scale, 1 = strongly disagree and 7 = strongly agree for each option in this statement). Consequences anti-shirking adds the scores, where the score on "The colleague resented it" is reversed.

Co-worker relations: Addition of the score on the following three statement: "I am attached to my colleagues", "I am appreciated by my colleagues", and "I find the workatmosphere with colleagues enjoyable" (1-7 scale, 1 = strongly disagree and 7 = strongly agree).

Employee-manager relations: "My store manager is inspiring and motivating" (1-7 scale, 1 = strongly disagree and 7 = strongly agree).

Job satisfaction: The job satisfaction variable is created by adding the scores from the following statements "My job is pleasant", "I am satisfied with my job", "I am actively searching for another job" (1-7 scale, 1 = strongly disagree and 7 = strongly agree), where the score on the last statement is reversed.

Observability effort: "It is easy to see whether colleagues are working hard or not." (1-7 scale, 1 = strongly disagree and 7 = strongly agree).

Chapter 5

Tournament Incentives in the Field: Gender Differences in the Workplace

Joint with Josse Delfgaauw, Robert Dur, and Willem Verbeke.

5.1 Introduction

Throughout the world, in business as well as in government, men are strongly over-represented in top positions. For instance, in 2008, only 16% of all ministerial positions worldwide were held by women; similarly, among the world's 192 heads of government, there were only eight women (IPU 2008). In business, the situation is not much different. For example, in a large sample of publicly traded US firms, Bertrand and Hallock (2001) find that only 2.5% of the five highest-paid positions are held by women. Wirth (2001) reports similar patterns for other countries.

Traditional explanations for the small number of women in top positions are occupational sorting resulting from gender differences in ability or preferences (Polachek 1981) and gender discrimination (e.g. Snizek and Neil 1992). Inspired by evolutionary biology, recent experimental studies – starting with Gneezy et al. (2003) – suggest a third explanation: men are more strongly motivated by competitive incentives or more effective in competitive environments than women, thus impeding

women in competitions for promotions or for new jobs.

By now, there is quite some empirical support for such gender differences. In a lab experiment, Gneezy et al. (2003) let participants solve computerized mazes and varied the competitiveness of the environment. They find that, while men and women perform equally well under individual piece rates, men perform much better than women under competitive incentives. Gneezy and Rustichini (2004) show that these gender differences are already present at a very young age. In a 40 meter dash, nine-year-old boys run much faster in a race than when they run alone. By contrast, while girls run as fast as boys when running alone, competition does not increase their running speed. In non-experimental settings, underperformance of women under competitive pressure is found in student admissions to schools (Jurajda and Münich 2008 and Örs et al. 2008) and in Grand Slam tennis (Paserman 2007). The recent field study by Lavy (2008), however, finds no gender differences in the effect of relative performance pay on high-school teacher's performance in Israel. Croson and Gneezy (2009) provide a recent overview of the literature.[1]

This chapterstudies gender differences in competition by conducting a field experiment in a naturally occurring work environment. A unique feature of our analysis is that we study competition among teams of employees, each headed by a professional manager. Using the variation in the gender composition of the teams as well as in the gender of the manager, we test whether female-dominated and female-led teams respond differently to competitive incentives, which were introduced in a random sample of the geographically dispersed teams. Moreover, we test for possible interaction effects between the gender of the manager and the gender composition of the team. Studying gender differences in competition among manager-led teams is most relevant in the context of the sharp gender differences in holding executive-level positions discussed above. Reaching an executive-level position, be it in business or government, commonly requires winning several promotion or job competitions.

[1] A closely related strand in the experimental literature studies self-selection into competitive environments. Datta Gupta et al. (2011), Dohmen and Falk (2011), Flory et al. (2010), and Niederle and Vesterlund (2007) find that men opt significantly more often for competitive compensation schemes than women. Gneezy et al. (2009) show that the reverse holds in a matrilineal society. Recent studies have shown that the gender gap in self-selection into competition by and large vanishes for girls attending single-sex schools (Booth and Nolen 2009) and when the tournament is among teams rather than among individuals (Dargnies 2009).

These competitions are often decided by candidates' relative performance which (except for employees at the lowest hierarchical level) depends not merely on one's own effort or talent, but also crucially on the performance of the members of the team one leads.

More concretely, we ran a field experiment in a discount retail chain in The Netherlands specializing in shoes, sports apparel, and casual clothing. About half of the 128 stores are led by a female manager, while across stores the percentage of female employees ranges from 50% to 100%. In a randomly selected subset of stores, we introduced short-term sales competitions among stores. The selected stores were divided into pools of 5 and competed for a period of 6 weeks on the basis of percentage sales growth compared to the same period the year before. All employees of the store with the highest sales growth over 6 weeks received a bonus of 75 euro; employees of each pool's runner-up received 35 euro. The stores that took part in a competition received weekly feedback in the form of a poster that ranked each store in their pool on their cumulative sales growth figures.

We find that, on average, the tournaments increase percentage sales growth by about five percentage points. We find no significant difference in the effect of tournaments on sales growth between stores with a male manager and stores with a female manager, nor do we find that sales competitions have a larger effect on performance in stores with a higher fraction of male employees. However, this masks a remarkable interaction effect of these two gender variables on sales growth responsiveness: in stores with a male manager, the effect of competition increases in the share of male employees, while the reverse holds for female-led stores. These effects are substantial.[2]

We can think of three plausible mechanisms behind this result. First, the response of team members to competition may crucially depend on the way a competition is communicated and promoted by the team's manager. Both male and

[2] Gneezy et al. (2003), Gneezy and Rustichini (2004), and Ivanova-Stenzel and Kuebler (2005) also study whether opponent's gender matters for performance under competition. We do not look into this issue, as teams have limited information, if any, on the gender composition of the stores they compete with. Casas-Arce and Martinez-Jerez (2009) analyze sales competitions among retailers organized by a commodities manufacturer. They do not study gender differences, but instead focus on the effect of the number of contestants in the tournament and on dynamic incentives.

female managers may have succeeded in making the competition appeal to employees of their own sex, but less so to employees of the opposite sex. Alternatively, the team nature of the incentive scheme may drive the difference in response. A male (female) manager may be better in strengthening the team's internal cohesion or curtailing free-rider problems if many team members are male (female). Lastly, as managers and employees were not randomised over stores, teams' gender composition may be the result of endogeneous matching on unobservables, which may correlate with teams' responsiveness to competitive incentives. We elaborate on these interpretations after presenting the results in Section 5.4.

A number of recent studies argue that competition can motivate people not merely because of the chance of winning a monetary reward, but also because of non-pecuniary benefits such as perceived esteem, status, and social recognition (Auriol and Renault 2008, Besley and Ghatak 2008, Frey and Neckermann 2008, Moldovanu et al. 2007). Kosfeld and Neckermann (2010) show experimentally that a tournament with no more at stake than an award of zero material value can have a great impact on people's performance. Likewise, Blanes i Vidal and Nossol (2009) and Azmat and Iriberri (2010) find that simply providing information to subjects about their relative performance boosts performance substantially. Bandiera et al. (2009), however, find the opposite effect. In our experiment, parallel to the treatment described above, another subset of stores competed in tournaments with the same setup except for the absence of a monetary reward for winning. So, stores also competed in pools of five, for a period of six weeks, and received a weekly ranking of stores in their pool based on sales growth, but neither the manager nor the employees could earn a bonus. We find that tournaments without monetary rewards have a significantly positive effect on sales growth. The effect is of similar magnitude as the effect of tournaments with monetary rewards, suggesting a high symbolic value of winning a tournament. Gender differences in the effects of competition are also similar in both treatments.

We proceed as follows. In the next section, we describe the experimental set-up and the data. Section 5.3 describes the methodology of our empirical analysis and Section 5.4 reports the results. Section 5.5 concludes.

5.2 Experimental set-up and data description

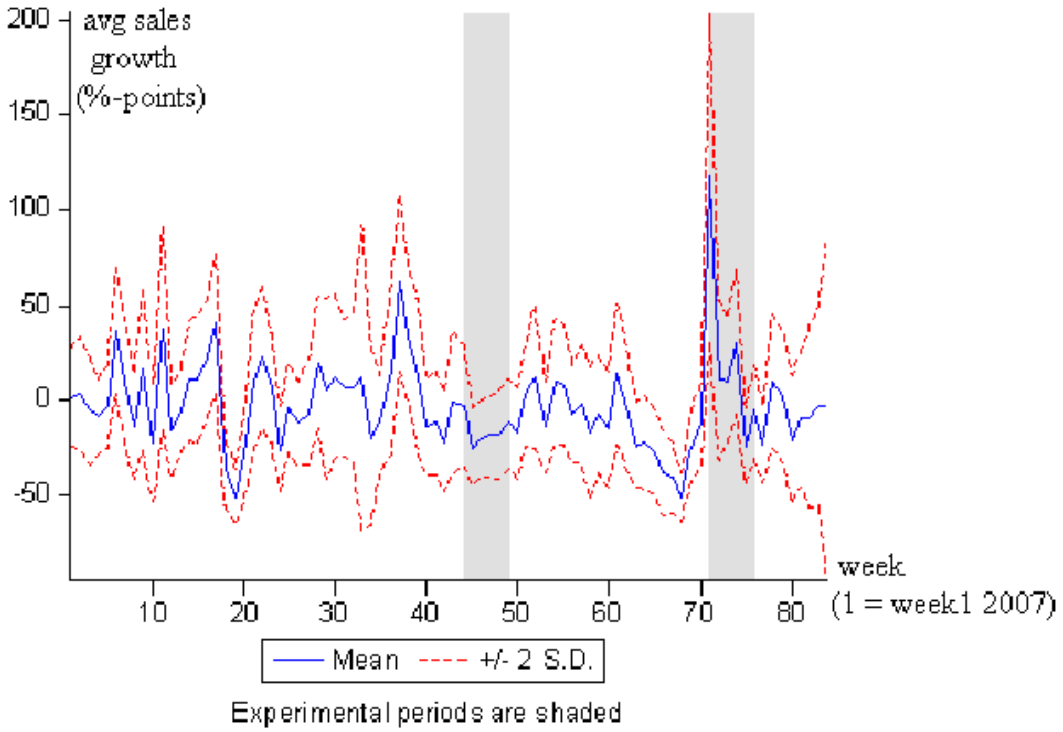
The field experiment took place in 2007-2008 in a discount retail chain in The Netherlands, selling male and female clothing, shoes, and sports apparel. The chain consists of 128 geographically dispersed stores operating under one brand name and employing a total of 1574 people. Store employees earn a flat hourly wage slightly above the legal minimum hourly wage. Store managers earn about 45% more and part of their pay is performance-related. On average, slightly less than 5% of a manager's earnings is performance-related.

The company's management wished to intensify the use of incentives. In consultation with the management, we designed sales competitions among subsets of stores. We used stores' percentage growth in sales as compared to sales in the same period a year earlier as the performance measure. Percentage sales growth is a commonly used performance measure in this company and is one of the key determinants of store managers' performance pay. We decided to introduce relative performance incentives rather than incentives based on absolute targets, as sales are very volatile (see Figure 5.1). A large part of this volatility is caused by common shocks (weather, holidays, advertising campaigns on national television, etc.),[3] which renders relative performance pay attractive (Lazear and Rosen 1981, Green and Stokey 1983, and Nalebuff and Stiglitz 1983).

In the sales competitions, stores competed in pools of five during a period of six weeks. Stores received weekly feedback in the form of a poster containing cumulative sales growth figures for all five stores in their pool, ranked in descending order. Store managers were instructed to put up these posters in the store's canteen, where employees drink coffee and have lunch. The posters as well as the instructions were sent to the store managers through the company's usual channels; store managers and store employees did not know they took part in an experiment. Hence, our experiment can be classified as a natural field experiment (Harrison and List 2004). Store employees were not informed about the sales competitions by the company's

[3] A panel regression including only week fixed effects explains about 65% of the variation in stores' sales growth. The spike in week 71 in Figure 5.1 is a common shock, most likely resulting from weather conditions; details are available on request.

Figure 5.1: Weekly percentage sales growth compared to the same week the year before



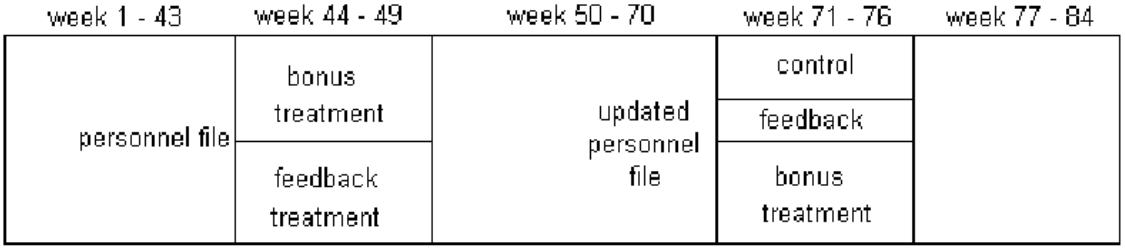
management; it was up to the store managers to promote the competition.

Our study comprises two experimental treatments and an untreated control. First, in the 'bonus' treatment, stores compete for a monetary reward. The store manager and all employees of the winning store received a reward of 75 euro; the manager and employees of the runner-up received 35 euro.[4] Second, in the 'feedback' treatment, no monetary reward could be won. Apart from the presence or absence of a monetary reward, the bonus treatment and the feedback treatment were identical.

Our dataset covers a period of 84 weeks (starting in week 1 of 2007). Sales competitions took place in two experimental periods of 6 weeks (in weeks 44 - 49 and weeks 71 - 76). Figure 5.2 gives an overview of all the events related to the experiment. In the first experimental round, all stores were assigned to one of the two experimental treatments, either bonus or feedback. In the second round, we

[4]The first prize was about 5 percent of an employee's monthly wage. Rewards were halved for part-time employees.

Figure 5.2: Overview of the experimental set-up and timing of events



included a control group of stores not taking part in a competition. We decided against a control group in the first round, because at that time we intended to focus our study on the effects of monetary rewards in tournaments. The second round gives us the opportunity to also assess the effects of tournaments per se.

Competition provides stronger incentives when contestants are more homogeneous (Lazear and Rosen 1981). Bearing this in mind, we used data on past sales performance to create relatively homogeneous pools of stores. The assignment procedure for the first round of the experiment was as follows. All stores were ranked according to percentage sales growth over the weeks 1 up to 37 compared to the same period the year before. The five stores with the highest sales growth were grouped into one pool and assigned to the bonus treatment; the next five were grouped into the next pool and assigned to the feedback treatment. This process was iterated consecutively until all 125 stores were grouped into 13 bonus treatment pools and 12 feedback treatment pools.[5]

The assignment procedure for the second experimental round was partly imposed by the company. For fairness reasons, the company obliged us to assign all stores who were in the feedback treatment during the first period to the bonus treatment in the second period.[6] We grouped these stores into new pools of five stores each, this time using sales performance in weeks 50 to 68 to create relatively homogeneous pools. The remaining stores were assigned either to the feedback condition or to the

[5]During the first round of the experiment, 3 stores were closed for renovation.

[6]The company wished, at a later point in time, to evaluate the experiment together with the store managers and feared that it would be considered unfair when some stores had never been assigned to the bonus condition.

Table 5.1: Descriptive statistics

			Second period assignment					
	All stores		Bonus treatment		Feedback treatment		Control group	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Weekly sales growth (percentage points)	-2.44	29.87	-2.62	29.21	-2.22	30.90	-2.32	30.09
Female manager[a]	0.44		0.48		0.36		0.43	
Age manager[a]	38.7	9.3	39.2	10.3	38.2	7.6	38.1	9.1
Tenure manager[a]	11.9	8.7	12.9	9.3	11.7	8.7	10.7	7.8
Number of employees	11.8	3.2	11.8	3.4	11.2	2.7	12.2	3.5
Percentage of female employees	84.9		85.1		81.9		84.5	
Average age employees	25.2	4.0	26.0	4.5	24.7	3.3	24.4	3.3
Renovated store	0.13		0.17		0.07		0.13	
Number of stores	128		60		30		38	

Figures are the averages of the values in the two personnel files (see Figure 5.2).

F-tests show that none of the differences in means between the two treatment groups and the control group is statistically significant at the 10% level.

[a] The first personnel file contained information on 122 store managers, and the second personnel file had information on 119 store managers; see footnote 7. The missing observations are not correlated with either of the treatments.

untreated control group according to a similar procedure as before: the five best-performing stores in weeks 50 to 68 were assigned to the control condition, the next five stores were grouped into a pool and became part of the feedback condition, and so on. To avoid confusion and diminish sabotage opportunities, we replaced a store when two stores from the same city happened to be assigned to the same pool. In both periods, we made two of these adjustments.

The company provided us with the weekly sales data of each store, presented in indexes for confidentiality reasons. We used these to calculate the percentage growth in sales as compared to sales in the same week a year earlier. We henceforth refer to this measure as weekly sales growth. We also received each store's personnel file before both experimental rounds, with information on gender, age, and tenure of the store's manager and employees. Descriptive statics are given in Table 5.1.Across all stores, average weekly sales growth was negative in the period we consider. The retail chain had slightly less female-led stores than male-led stores.[7] The average

[7]In both personnel files, for some stores information about the manager is missing, either because the store temporarily had no manager, or (in a single case) a store had two managers. In one store, a male manager was replaced by a female manager in between the two experimental

store had 12 employees (excluding the store manager), of which 85% was female. The average age and tenure of store managers was 39 years and 12 years, respectively. Some of the stores underwent a renovation, which made their appearance more modern without changing the range of products sold. Before the first experiment, six stores had been renovated; at the start of the second experimental period, an additional 9 stores had been renovated. As stores are closed during renovation, there are 268 missing store-week observations. In the analysis we control for the effects of renovation on subsequent sales growth by including a dummy variable which takes value 1 from the week in which a renovated store is reopened onwards.

Table 5.1 also reports the descriptive statistics within the two treatment groups and the control group to which stores were assigned in the second experimental period (see Figure 5.2). The three groups of stores hardly differ on observables. A randomization check using F-tests reveals that there are no statistically significant differences in the means of the observables between the three groups. Table 5.2 reports the descriptive statistics separated by store managers' gender, where we only include the 114 stores we use in analyzing the gender effects (see footnote 7). Over the whole period, male-managed stores reached 0.4 percentage points higher weekly sales growth than female-managed stores, but the difference is not statistically significant. Male managers have significantly longer tenure than female managers, and run stores with significantly more employees. In the analysis below, we perform robustness checks where we control for these differences. Importantly, there is quite a lot of variation in the percentage of female employees, both for male- and female-led stores. Figure 5.3 depicts the distribution of the percentage of female employees in stores, separated by managers' gender. Gender of the manager is not significantly related to the gender composition of store employees. Note that there are no stores with a majority of male employees. This implies that our estimates of the effect of stores' gender composition are based on, and, hence, relevant for female-dominated teams.

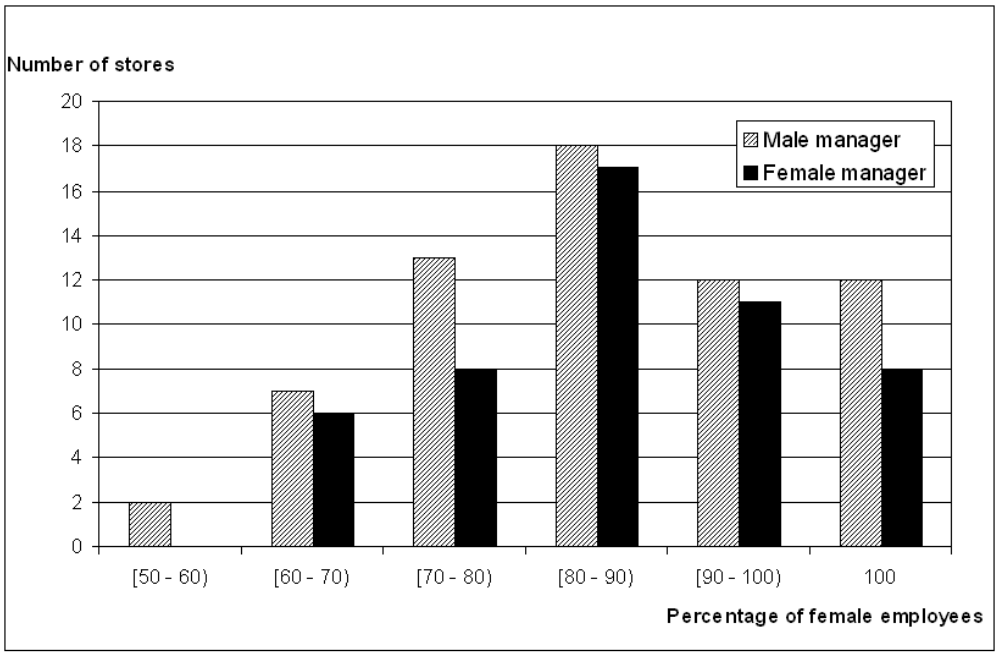
periods. When analysing gender effects, we exclude these stores from the analysis. This leaves 114 stores. In five other stores, the manager was replaced by a manager with the same gender; excluding these stores from the analysis does not affect the results qualitatively.

Table 5.2: Descriptive statistics by store managers' gender

	Gender manager			
	Male		Female	
	Mean	Std	Mean	Std
Weekly sales growth (percentage points)	-2.07	29.52	-2.46	30.79
Age	39.40	9.46	37.84	9.07
Tenure*	14.86	9.46	8.32	6.38
Number of employees*	12.60	3.21	10.71	2.99
Percentage of female employees	84.23		85.81	
Average age employees	25.96	3.56	24.62	3.74
Renovated store	0.17		0.10	
Number of stores	64		50	

This table includes only the 114 stores used in the analysis of the gender effects, see footnote 7.
Apart from sales growth, the figures are the averages of the values in the two personnel files.
* Difference in means for male and female managers statistically significant at the 5% level.

Figure 5.3: Distribution of the percentage of female employees by store managers' gender



5.3 Method

We estimate the effects of the competitions on sales growth using OLS panel estimation including week and store fixed effects. Let y_{st} be the sales growth of store s in week t. Further, let B^1_{st} be a dummy variable which is equal to one during the *first* experimental round when store s was assigned to the bonus treatment (rather than to the feedback treatment). Similarly, let B^2_{st} and F^2_{st} be dummy variables for whether in the *second* experimental round, store s was assigned to the bonus treatment and to the feedback treatment respectively (rather than to the control group). To assess the average effect of the treatments in both experimental periods, we estimate:

$$y_{st} = \alpha_s + \theta_t + \gamma B^1_{st} + \delta B^2_{st} + \mu F^2_{st} + \upsilon R_{st} + \varepsilon_{st} \tag{5.1}$$

where α_s and θ_t are store fixed effects and week fixed effects, respectively, R_{st} is a dummy for whether store s had been renovated before week t, and ε_{st} is an error term. We cluster standard errors at the store level to correct for serial correlation within stores as well as for heteroscedasticity across stores (see Bertrand et al. 2004 for a discussion of the importance of correcting for serial correlation in differences-in-differences estimation).

Observe that we allow the effect of the bonus treatment relative to the feedback treatment to differ between the first and second experimental round, i.e., we do not restrict that $\gamma = \delta - \mu$. Loosely speaking, for each experimental period, we estimate differences-in-differences effects of the treatments, where we assume that in all non-experimental weeks, it is 'business-as-usual' for all stores. Hence, we do not allow for carry-over effects of treatments into the weeks following an experimental period. We have checked the robustness of this approach in two ways. First, none of our results is affected qualitatively if we exclude the first 8 weeks after either experimental period (weeks 50 - 57 and/or weeks 77 - 84) from our analysis. Second, all our results from the first experimental period carry over to an estimation which includes only the first 49 weeks (i.e., which excludes all weeks after the first period; see Figure 5.2). Similarly, we find qualitatively similar results for the treatment effects in the second experimental period if we include only the period after the first experimental period

(week 50 onwards).[8]

Besides the average treatment effects, we investigate how these treatment effects depend on the gender of the store manager, the gender composition of the store's employees, the interaction between these two, and the store's team size. To study these issues, we add interaction effects to equation (5.1). Let E_t^1 be a dummy that takes value one for all observations in the first experimental period (weeks 44 - 49) and, similarly, let E_t^2 be a dummy that takes value one in the second experimental period (weeks 71 - 76). The effect of variable X_s^p on the effect of our treatments is estimated by:

$$y_{st} = \alpha_s + \theta_t + \gamma B_{st}^1 + \eta X_s^1 B_{st}^1 + \kappa X_s^1 E_t^1 + \delta B_{st}^2 + \mu F_{st}^2 + \lambda X_s^2 B_{st}^2 + \pi X_s^2 F_{st}^2 + \psi X_s^2 E_t^2 + \upsilon R_{st} + \varepsilon_{st} \tag{5.2}$$

where X_s^p is the value of the variable X for store s in experimental period $p \in \{1, 2\}$ as taken from the personnel files received just before period p.[9] Again, we allow for differences between the estimated effect of variable X_s^p on the effect of the bonus treatment relative to the feedback treatment between the first and second experimental period (η versus $\lambda - \pi$). The inclusion of the E_t^1 and E_t^2 terms is necessary to obtain differences-in-differences estimates of the effect of X_s^p on the treatment effects. Not including these terms would imply that period-specific effects of X_s^p on stores' sales growth would be erroneously picked up by the interaction effects with the treatments (i.e., by η, λ, and π).

[8] We cannot, however, identify any possible carry-over effects of first-period assignment to second-period treatment effects, as assignment in the second period was not completely random (see Section 5.2). Note, however, the long time lag between the first and the second round (20 weeks).

[9] Obviously, in most of our regressions X_s^p is a vector of variables and, likewise, η, κ, λ, π, and ψ are vectors of coefficients. There is not enough variation in X_s across periods to include these variables as separate regressors. However, the effects of these lower-level terms of the interaction are captured by the store fixed effects. As a robustness check, we also estimated a random effects regression model, where we took up X_s as separate regressors. We find qualitatively similar results to those reported in the next section. Details are available upon request.

5.4 Results

The first column in Table 5.3 gives the results of estimating (5.1). Focusing on the second round, we find that the bonus treatment and the feedback treatment both have positive average treatment effects on weekly sales growth. This average treatment effect is statistically significant for the feedback treatment, and is borderline significant for the bonus treatment with a p-value of 0.13.[10] The size of the effects is also economically significant, as stores in the bonus and feedback treatment achieve 4.8 and 6.9 percentage points additional sales growth, respectively.[11] The difference in the effects of the bonus treatment and feedback treatment in the second period is not significant: a Wald test on the restriction that the effects are equal ($\delta = \mu$) has a p-value of 0.32. In the first round of the experiment, stores in the bonus treatment perform slightly better than stores in the feedback treatment, however this difference is again insignificant.[12] Hence, the financial reward in the bonus treatment did not trigger additional sales growth on top of the effect of the tournament that is also present in the feedback treatment.

Next, we allow the treatment effects to vary by the gender of the store manager and by the gender composition of stores' employees. Since we did not find any difference between the two treatments, we pool the observations of both competitions; i.e., we investigate gender differences in the response to competition irrespective of whether a monetary prize could be won.[13] Pooling the observations gives a slightly different specification:

$$y_{st} = \alpha_s + \theta_t + \delta(B_{st}^2 + F_{st}^2) + \lambda X_s^2(B_{st}^2 + F_{st}^2) + \psi X_s^2 E_t^2 + \upsilon R_{st} + \varepsilon_{st}. \qquad (5.3)$$

When we estimate (5.3) with X_s^2 only including a female manager dummy, we find

[10] Throughout the thesis, p-values are based on two-sided tests.

[11] Lack of data on the absolute value of sales and profit margins implies that we cannot establish whether this increase in sales outweighs the cost of the tournaments. However, the company's management was content with these results.

[12] Furthermore, we cannot reject the hypothesis that $\gamma = \delta - \mu$ in (5.1), i.e., that the differences between the effects of the bonus and the feedback treatments are equal in the first and second round (p-value is 0.18).

[13] The results are by and large the same when we do not pool the treatments, as can be seen by comparing the first column of Table 5.4 with the second column of Table 5.3, which reports the results of estimating equation (5.2).

Table 5.3: The effect of competition on sales growth

	Dependent variable: Sales growth			
Independent variables	(1)		(2)	
First round: **Base category = Feedback treatment**				
Bonus treatment	0.995	(0.986)	1.763	(1.667)
Female Manager x Bonus treatment			-1.590	(2.075)
Female Manager x Experimental round 1			-1.057	(1.292)
% Female employees x Bonus treatment			-0.058	(0.119)
% Female employees x Experimental round 1			0.168	(0.078)**
Female Manager x % Female employees x Bonus treatment			0.172	(0.155)
Female Manager x % Female employees x Experimental round 1			-0.187	(0.100)*
Second round: **Base category = Control group**				
Bonus treatment	4.796	(3.133)	3.537	(3.491)
Feedback treatment	6.902	(3.494)**	3.346	(3.290)
Female Manager x Bonus treatment			-1.832	(4.633)
Female Manager x Feedback treatment			4.098	(5.968)
Female Manager x Experimental round 2			3.234	(3.563)
% Female employees x Bonus treatment			-0.384	(0.261)
% Female employees x Feedback treatment			-0.416	(0.229)*
% Female employees x Experimental round 2			0.413	(0.170)**
Female Manager x % Female employees x Bonus treatment			0.831	(0.347)**
Female Manager x % Female employees x Feedback treatment			1.019	(0.463)**
Female Manager x % Female employees x Experimental round 2			-0.547	(0.231)**
Renovated store	8.315	(3.554)**	7.866	(3,781)**
Period fixed effects	YES		YES	
Store fixed effects	YES		YES	
Number of stores included	128		114	
Observations	10484		9326	
R^2	0.693		0.705	
Log likelihood	-44292.3		-39284.8	

***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

% Female employees is mean-centered.

Standard errors clustered at store level in parentheses.

only insignificant interaction effects (not reported for brevity). Hence, on average, the effect of competition on sales growth does not differ between stores with a male manager and those with a female manager. Similarly, we find insignificant interaction effects if X_s^2 only includes the percentage of female employees in a store. Thus, across all stores, we find no evidence that the gender composition of store employees influences the effects of competition. However, interacting the store manager's gender and the gender composition of store employees reveals an interesting pattern. The first column of Table 5.4 gives the results of estimating (5.3), where the treatments are interacted with both a female manager dummy and the percentage of female employees, as well as interacted with the interaction between the female manager dummy and the percentage of female employees. Thus, we allow for different effects of the gender composition of the stores' personnel on the effect of competition in stores with a male manager compared to stores with a female manager.[14]

Remarkably, we find that the sign of the effect of the percentage of female employees on a store's responsiveness to competition depends on the gender of the store manager. In stores with a male manager, the effect of competition on sales growth significantly decreases in the share of female employees with a marginal effect of -0.404 percentage point sales growth. An increase of one standard deviation in the percentage of women employed in a store, or about 12.5 percent points, leads to a decrease of about 5 percentage points in the treatment effect for male-led stores. By contrast, in female-led stores the responsiveness to competition increases in the percentage of female employees, with a marginal effect of $-0.404 + 0.886 = 0.482$ percentage point sales growth. A Wald test shows that this effect differs significantly from zero (p-value is 0.03). The magnitude of this effect is the same as in male-led stores: an increase of one standard deviation in the percentage of female employees increases the effect of competition by 6 percentage points.

The estimated treatment effects for various manager/employee combinations are depicted in Figures 5.4 and 5.5. Figure 5.4 gives the point estimates and the 95% confidence intervals of the effect of competition in the second round for male-managed

[14] In the estimation reported in Table 5.4, the percentage of female employees is mean-centered. Table 5.4 only reports the coefficients from the second round of the experiment, as we did not include a control group in the first round of the experiment, see Section 5.2.

Table 5.4: The effect of competition on sales growth: gender differences

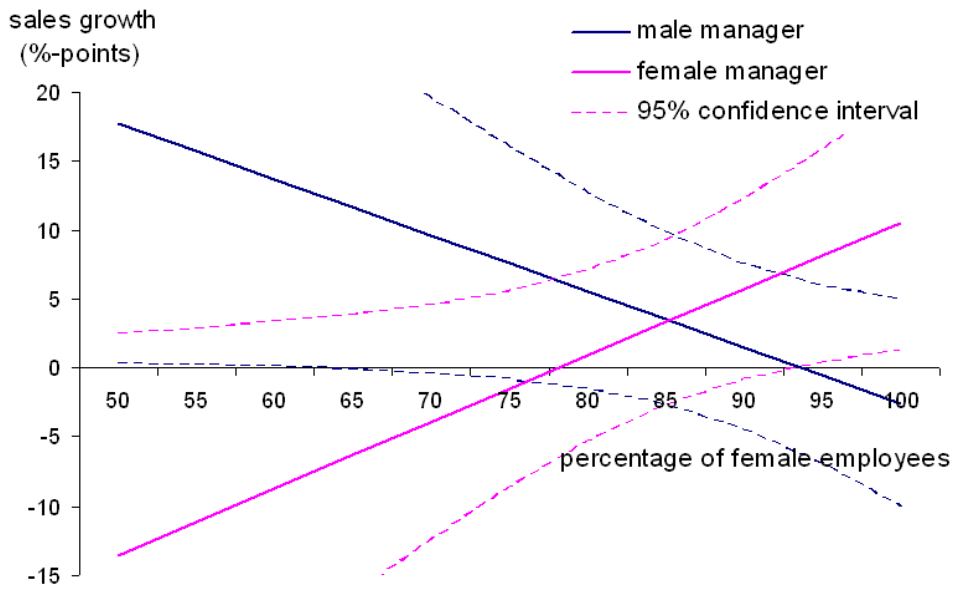
	Dependent variable: Sales growth			
Independent variables	(1)		(2)	
Second round: **Base category = Control group**				
Competition (=Bonus+Feedback)	3.555	(3.183)	3.995	(3.779)
Female Manager x Competition	-0.206	(4.391)	-0.964	(5.346)
Female Manager x Experimental round 2	3.411	(3.551)	4.021	(4.751)
% Female employees x Competition	-0.404	(0.212)*	-0.413	(0.213)*
% Female employees x Experimental round 2	0.408	(0.169)**	0.413	(0.170)**
Female Manager x % Female employees x Competition	0.886	(0.304)***	0.867	(0.300)***
Female Manager x % Female employees x Experimental round 2	-0.548	(0.230)**	-0.533	(0.224)**
Number of employees x Competition			-0.351	(0.754)
Number of employees x Experimental round 2			0.255	(0.706)
Renovated store	7.750	(3.760)**	7.774	(3.862)**
Period fixed effects	YES		YES	
Store fixed effects	YES		YES	
Number of stores included	114		114	
Observations	9326		9326	
R^2	0.705		0.705	
Log likelihood	-39290.4		-39290.1	

***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

% Female employees is mean-centered.

Standard errors clustered at stores level in parentheses.

Figure 5.4: Gender differences in the effect of competition on sales growth depicted with a 95 percent confidence interval of the estimates

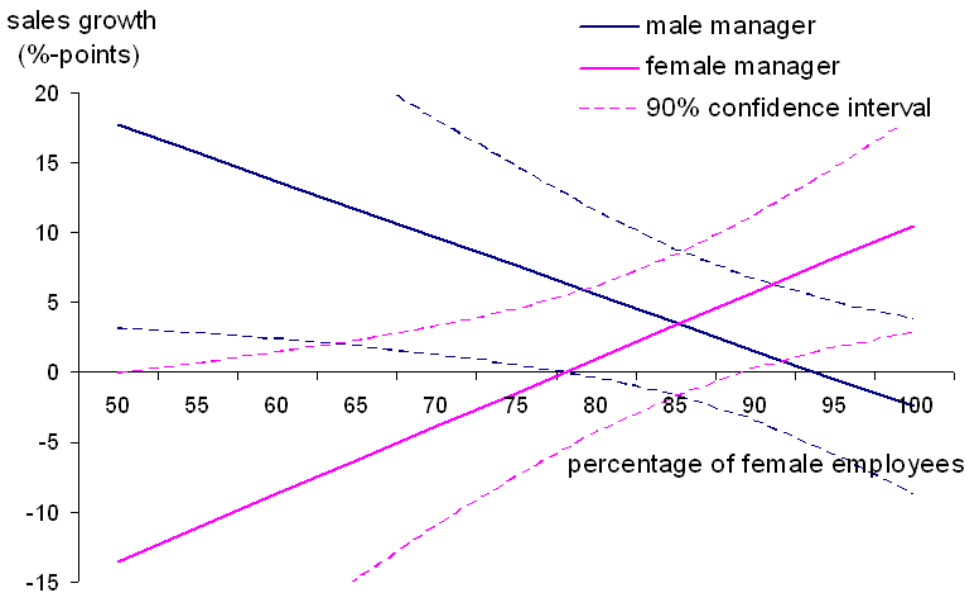


and female-managed stores separately. Figure 5.5 depicts the same but with 90% confidence intervals. Both figures clearly show that the competition has been most effective in raising sales growth in male-led stores with a relatively high percentage of male employees. Figure 5.5 shows that the effect of competition is statistically significant in male-managed stores as long as the percentage of women employed does not exceed 80%. In female-managed stores, the pattern is reversed: the estimated impact of competition strongly increases with the percentage of female employees. The effect of competition in these stores is significant when at least 90% of the employees is female.

We have checked the robustness of our findings by controlling for several other variables. First, the results reported in Table 5.4 are not affected if we control for managers' tenure or for managers' age. Similarly, neither employees' average age nor average tenure in a store affects our results. Lastly, none of our results is affected qualitatively when weighing employees by their full-time equivalent.

Overall, our findings give a nuanced picture of gender differences in manager-led team performance under competition. It is not gender per se that affects perfor-

Figure 5.5: Gender differences in the effect of competition on sales growth depicted with a 90 percent confidence interval of the estimates



mance under competition, but rather the match between the team's manager and the gender composition of the team: competition positively affects performance when the manager and a sufficiently high percentage of employees have the same gender. As mentioned in the Introduction, we can think of three plausible mechanisms behind this result. First, the response of team members to competition may crucially depend on the way a competition is communicated and promoted by the team's manager. In our experiments, we deliberately left a lot of discretion to team managers on how to use the competition as an incentive device. In particular, both the announcement of the competitions and the weekly posters were only sent to the store managers, not to the employees. It was up to the team managers to make the competitions appealing to their employees. Managers may have succeeded in making the competitions appealing to team members of their own sex, but less so to team members of the opposite sex.[15] This interpretation is well in line with evidence from

[15]The idea that dissimilarity in personal attributes such as gender can deteriorate communication in organizations dates back to at least March and Simon (1958), who argue that dissimilarity may give rise to 'language incompatibility' and less frequent communication. There is pervasive evidence for 'homophily': the tendency that people interact more frequently with people with similar

management studies showing that when working for a manager of the opposite sex, employees find their duties and responsibilities much more ambiguous than when working for a manager of the same sex (Tsui and O'Reilly 1989, McNeilly and Russ 2000). Relatedly, a number of studies in organizational psychology have shown that, as compared to employees with opposite-sex managers, those with same-sex managers are more likely to develop high-quality leader-member exchange relationships (LMX) – a widely used measure of manager-employee mutual support, trust, and obligation – which may in turn facilitate communication.[16] Lastly, experimental evidence using a subject pool of both students and banking executives finds that female participants tend to feel more comfortable supervising a female person than a male person in a challenging task, while male participants expect fewer conflicts with a male subordinate and perceive males to be more competent in a challenging task (Mai-Dalton and Sullivan 1981).

Alternatively, as team composition was not randomized in our experiment, the teams' gender composition may be the result of endogeneous matching on unobservables. If these unobservable characteristics are correlated with teams' responsiveness to competition, the pattern we find may arise without there being a causal link between teams' gender composition and performance under competition. For instance, suppose that relatively competitive managers have a preference for supervising employees of their own gender. These managers will self-select into (or gather) a team with more employees of their own gender than less competitive managers, and they will respond more strongly to tournament incentives. In this scenario, our results are driven by unobserved managerial characteristics rather than by team composition. Our experimental design does not allow us to discriminate between these mechanisms, so that this remains an important open question we hope to address in future work.

A third possible mechanism behind our results might be that male managers

rather than dissimilar characteristics (McPherson et al. 2001, Borgatti and Foster 2003, Reagans 2005). While we find strong effects of gender similarity, we find no effect of similarity in age or tenure on stores' responsiveness to competition. This suggests that it is indeed gender similarity rather than similarity in general that drives our results.

[16] See e.g. Duchon et al. (1986), Pelled and Xin (2000), and Varma and Stroh (2001). Wayne et al. (1994) discuss a number of reasons for why these differences may arise.

are better at reducing free-rider behavior in a team with many male employees, and likewise for female managers with female-dominated teams. We address free-rider behaviour by using the number of workers employed in the store, which varies between 5 and 20 employees across stores. Note, though, that we do not have an ideal set-up to analyze free-rider effects, as store size is not randomized. When there is an unobserved, systematic difference between small and large stores that affects the responsiveness to competitive incentives, this is reflected in the estimates, so that these are to be interpreted with caution.[17] The second column of Table 5.4 gives the results of including the number of employees interacted with the competition dummy, i.e., the pooled treatments. We find no indication for free-riding behavior. The effect of competition does not depend on the number of employees.[18] Also, the inclusion of the number of employees hardly affects the estimates of the gender effects on performance under competition. Finally, interacting the number of employees with the gender composition of stores yields results that do not support the interpretation that free-rider behavior is reduced in stores where the manager and a large part of the store's employees are of the same gender; details are available upon request.

5.5 Concluding remarks

We have studied how teams led by a professional manager respond to competitive incentives. Overall, we find strong effects, even when there is no monetary reward to winning the competition. Further, our results suggest that the gender of the manager and the gender composition of the team jointly affect performance under competition. Male-led teams are more responsive to competition when a larger fraction of the team members is male. By contrast, female-led teams respond more strongly to competition when the fraction of female members is larger. If generalizable, our results give rise to some optimism about future reductions in the stark gender inequalities in executive positions in business and government that we discussed in

[17] For instance, free-rider effects are mitigated when managers with better team-building capabilities are more likely to be assigned to larger stores.

[18] The same conclusion is drawn when the gender interaction terms are excluded from the estimation (so that X_s^2 in (3) only includes the number of employees).

the Introduction. As women have massively entered the labor market over the last decades, work teams have become more gender-diverse. Our results suggest that this should put female managers on a more equal footing in contests for promotions.

Chapter 6

Summary and directions for further research

In this thesis, I study how organizations can incentivize a group of individuals, both in theory and by conducting field experiments. The field experiments introduce short-term team incentives in a randomly selected sample of stores that belong to a single Dutch retail chain of 128 stores. The results offer valuable lessons for both researchers and practitioners alike. In addition, this thesis provides a theoretical investigation of the interaction between the remuneration schemes that organizations offer to its employees and the interpersonal relations between colleagues. Hereby, this thesis gives insights into an underexposed aspect of teamwork: The social relationships between colleagues, which is a job aspect that is highly valued by most people. This chapter provides a summary of the findings, discusses some policy implications, and makes suggestions for further research.

6.1 Summary

Chapter 2 identifies a channel through which financial incentives for productive activities can affect social interaction between colleagues. We develop a principal-multi-agent model where agents do not only choose their productive activities, but also their social interaction with colleagues, which in turn creates co-worker altruism. We show that, in the absence of team or relative performance incentives, workers

do not invest enough in their relationships with co-workers, as the benefits from relationship-building are not fully internalized. This externality problem comes at a cost to the employer, as social interaction with colleagues is a valued job attribute: Good co-worker relationships allow employers to attract and retain workers for lower wages. Employers can stimulate social interaction among colleagues by providing either team incentives or relative performance incentives. These incentives give workers a strategic motivation to invest in the relationships with their co-workers, as the externality problem in effort provision is less severe with a more altruistic colleague. Incentives for effort are restored through fine-tuning of individual incentives.

The predictions can be distinguished from those developed in related literature. We predict that, under the assumption of conditionally altruistic or reciprocal agents, team incentives and relative performance incentives lead to more social interaction and better co-worker relations. Instead, theories of peer pressure in teams or sabotage in tournaments predict worse co-worker relations in response to those incentives (Kandel and Lazear 1992, Lazear 1989). Theories of helping behavior need some complementarity in productive efforts to come to the same prediction (FitzRoy and Kraft 1986). Finally, a team bonus on top of flat wages does not lead to better co-worker relations in models that assume inequality-averse agents (Bartling 2011). Next to the potential positive effect of team and relative performance incentives for co-worker relations, a second managerial implication is that social relations matter for the effects of incentives. If a manager is unaware of the social setting and does not fine-tune individual incentives for effort, then team bonuses may give too strong incentives, while relative performance incentives may be too weak.

Chapter 3 also studies how interpersonal relationships between colleagues may influence the optimal use of incentives. I study a setting where co-workers have the best information about each other's effort. Managers may try to retrieve and exploit this information through peer evaluation. Chapter 3 studies peer evaluation in a pure moral hazard model of production by two limitedly liable agents. Agents receive signals about their colleague's effort level, and are asked to report them to the principal. An individual bonus for the receipt of a positive evaluation stimulates effort as long as signals are revealed truthfully, because agents desire to increase

the likelihood that their colleague receives a positive signal and rates performance accordingly. The agent's evaluation decision is characterized by a trade-off between a cost of lying about the received signal and the internalized benefit (or cost) from a bonus to a friend (or foe). Lying costs ascertain that there can be truthful revelation of signals, while interpersonal relations may cause a bias. Nevertheless, the principal always includes a bonus for receiving a positive peer evaluation in the optimal contract. The principal will only use a team bonus in addition to the peer evaluation bonus in case the latter becomes severely constrained by strong social relations between colleagues or small lying costs.

The result that co-worker relations, either good or bad, constrain the peer evaluation bonus so as to keep evaluations truthful has a testable implication: Incentive effects of peer evaluation diminish with more pronounced social relations between co-workers. This result also has an implication for management practice, as managers should organize peer evaluations for employees that are rather indifferent towards one another. The chapter contributes to an ongoing debate in the management literature about whether peer evaluations can be used to determine merit pay and promotions, or whether they should only be used for training or development (see among others Edwards and Ewen 1996, Coates 1998). We showed that the concern of a (dis)likability bias can be overcome by using a smaller bonus for peer evaluation. Moreover, the chapter illustrated that even invalid peer evaluations can have incentive effects through the influence over a colleague's expected costs of lying; e.g., employees don't want to put a befriended colleague in the difficult position of having to lie about their performance.

Chapter 4 describes the results from a field experiment on the determinants and effects of anti-shirking behavior in a retail chain. Stores are randomly assigned to a temporary team incentive treatment or a control group. The team incentive does not increase sales performance during the experiment. The team bonus has a negative effect on employees' stated willingness to undertake anti-shirking behavior. We present suggestive evidence for the claim that employees in treated stores unsuccessfully attempted to raise group performance by exerting more peer pressure during the experiment, at the expense of co-worker relations. Questionnaire results

revealed that employees in treated stores stated more often that colleagues reacted negatively in response to anti-shirking behavior, less often that colleagues improved their behavior, and were less positive about the social relations with co-workers.

Related studies have shown that peer effects can be productive (Mas and Moretti 2009 and Falk and Ichino 2006). Therefore, organizations should strive to create and nurture a strong anti-shirking culture. In line with Freeman et al. (2010) we identify a possible instrument to stimulate the willingness to undertake anti-shirking behavior; namely, management attention, as employee-management relations have a positive influence on anti-shirking behavior. Further, our study extends the analysis in Freeman et al. (2010) by taking advantage of our experimental set-up. The chapter shows that, in contrast to Freeman et al. (2010), the team incentive has a negative effect on the anti-shirking culture. Whether the team incentives has a negative impact on how colleagues interpret anti-shirking behavior or whether the negative effect is only present when increased attempts to reduce shirking are ineffective remains unclear from our study.

Chapter 5 presents the results of a field experiment that introduces short-term sales competitions among random subsets of stores. We introduce two experimental treatments, which were identical apart from the presence or absence of a monetary reward for the winning store and runner-up. Remarkably, sales competitions with and without monetary rewards have a similar positive effect on sales growth, suggesting a high symbolic value of winning a tournament. Further, in contrast to recent literature on gender differences in competition (Gneezy et al. 2003, among others), we find that the positive effect does not vary by the gender of the store manager nor by the gender composition of stores' employees. However, this zero-result masks an interesting interaction effect of these two gender variables for the responsiveness to competition: The sales competitions only have a large effect on sales growth in stores where the store's manager and a sufficiently large fraction of the employees are of the same gender.

A positive effect of sales contests without financial rewards may be surprising to mainstream economics, but strong effects of non-monetary rewards of winning a contest have been documented before (Kosfeld and Neckermann 2011). This result

offers a valuable insight to organizations, as there is the potential to motivate employees without incurring high costs. Some of these possibilities are presented in the economic analysis of awards (Frey and Neckermann 2008). Second, we contributed to the literature on gender differences by showing that competition can be equally stimulating to men and women. Gender differences in competition were also absent in another field setting (Lavy 2008). Finally, Chapter 5 reveals an effect of interaction between the gender of a manager and the gender composition of the team. This interaction may be relevant for promotion tournaments between managers, as the performance of managers usually depends both on the effort of the team one leads and on one's own performance. This result suggest that an underrepresentation of women in the labor market makes it more difficult for female managers to reach top positions.

6.2 Directions for further research

Chapter 4 shows that co-worker relations suffer from the introduction of a temporary team incentive. This result is in line with the predictions of peer pressure theories and in stark contrast to the hypothesis we put forth in Chapter 2. However, in Chapter 2 we discuss empirical evidence of a positive relation between team incentives and co-worker relations. In the light of these seemingly contradicting findings, there is a need for a unifying framework that identifies under which circumstances we can expect peer pressure or social interaction to prevail. In addition to combining different actions into one framework, the analysis can benefit from introducing heterogeneity in workers' social preferences. Kosfeld and von Siemens (2009, 2011) make interesting steps in this direction, where different types of workers sort into firms offering different incentive schemes.

Another interesting next step to the analysis in Chapter 2 would be to study a situation where the employer only observes differences in performance or team output. In this case, team incentives and relative performance incentives will serve a dual role of promoting productive effort and stimulating social interaction. In such situations, it can become optimal for employers to discourage social interaction.

The team incentive may surpass one of its goals and as a side-effect of strong team incentives workers may become too concerned with pleasing each other. Likewise, high rewards in a promotion tournament may lead to excessive social interaction. In theory, it also seems possible that efficient incentives for effort are reached before social interaction reaches its first-best level, in which case the principal still desires to promote social interaction.

The behavioral assumption we made in Chapter 2, that social interaction affects altruism (i.e., conditional altruism), can also be applied to Chapter 3. In this case, it is likely that peer evaluations will have a positive influence on social interaction between agents. An agent may want to invest in the relationship with a colleague that evaluates his performance, so as to increase the likelihood of a positive evaluation by this colleague. By this logic the effectiveness of peer evaluation may deteriorate eventually. Other factors that could undermine the truthful peer evaluation are competition or collusion between agents that evaluate each other.

The result that a (dis)likability bias in peer evaluation can be overcome by smaller rewards may be sensitive to the assumption of the binary signal about performance. In a richer signal space, the trade-off between lying costs and the internalization of a colleague's utility from the peer evaluation bonus will probably lead to an optimal bias in the agent's evaluation message. My conjecture is that this poses no threat to the value of peer evaluation for the principal as long as he has complete information about the agents' preferences, as this allows him to infer how signals translate into evaluation messages. Otherwise, the principal may want to limit the number of evaluation messages that agents can send, such that messages provide him with information about the signal again.

The prediction that incentive effects of peer evaluation decrease with the intensity of social relations between co-workers is not easily tested in a field experiment, as one needs information on workers' preferences. Co-worker relations can be gauged with a questionnaire, but perhaps more promising is the combination of a field experiment with a laboratory experiment to elicit preferences (Falk and Heckman 2009). In this case, one could assign employees to a peer evaluation treatment, stratified by their social preferences and inhibitions towards lying.

In Chapter 4 we find that team incentives can harm the anti-shirking culture. This observation is in contrast to earlier findings, where anti-shirking behavior seems to be more prevalent in the presence of team incentives (Freeman et al. 2010). We cannot pin down the mechanism that caused a drop in the willingness to undertake anti-shirking behavior, as we do not observe behavior during the experiment. Our interpretation of the results is that attempts to raise group performance by exerting more peer pressure were ineffective, leading to a dissatisfaction with the consequences of their anti-shirking behavior and worse co-worker relations. However, we cannot distinguish our interpretation from some alternatives. Anti-shirking behavior may have become less appreciated in the presence of the team incentive, as colleagues can suspect that anti-shirking behavior is inspired by greed. Further research on the mechanisms that drive anti-shirking behavior in the presence of team incentives is needed.

Chapter 5 makes a contribution to a recent literature on gender differences in competition. Before I point out some further research ideas, let me provide the reader with a critical side-note to the gender differences literature. Some scholars from psychology advocate caution with research on gender differences, or to abandon this research completely, as results could be used to perpetuate stereotyping or justify discrimination and oppression (Baumeister 1988, Eagly 1995). Further, if one observes gender differences, they may be the result of a mix of other variables such as social norms, hormones, or physical stature, on which Baumeister states: "All of these variables are worth study, but studying sex differences is a poor substitute for studying them, just as studying racial differences is a poor substitute for studying socioeconomic class, as some sociologists have done." (p. 1094) However, insights into gender differences, from whatever sociological or biological process they originate, can potentially offer explanations for stylized facts that are otherwise easily ascribed to discrimination.

Our results suggest that the gender difference in the response to competition found in laboratory experiments may be absent in the workplace. However, the possibility exists that the company in which our experiment took place employs a selection of more competitive women, or men with more feminine tendencies.

There is a need for more field experiments on gender differences in competition to conclude whether the findings in the lab generalize to the workplace. Further, we give a tentative explanation for our finding that the effect of competition depends on interaction between the manager's gender and the gender composition of a team: A manager may be more able to make the tournament appealing to team members of the same gender. However, an alternative explanation could be that our finding results from endogenous matching on unobservables, as managers and employees were not randomized over stores. In case these unobservable characteristics also influence the responsiveness to competition, our finding may arise without a causal link between gender composition and the response to competition. Lab experiments are arguably the only method that allow for a cleaner replication of this part of our study, as random assignment of managers and employees to teams is almost impossible to achieve in the field, especially in established organizations.

Samenvatting
(Summary in Dutch)

Introductie

Veelal bestaat er een belangenconflict tussen werknemers en werkgevers over de hoeveelheid werk die verricht dient te worden. Dit belangenconflict kan leiden tot een probleem van 'moral hazard' in het geval dat de gevraagde werkzaamheden onmogelijk of zeer kostbaar te achterhalen zijn.[1] Organisaties moeten ervoor zorgen dat de werknemer in eigen belang de doelen van de organisatie nastreeft. Een veel bestudeerde wijze om dit te realiseren is door middel van het inkomen van de werknemer afhankelijk te maken van de uitkomsten; bijv. stukloon, een jaarlijkse bonus of promotie voor buitengewone prestaties. De positieve invloed van deze financiële prikkels op de productiviteit is bewezen in verschillende omgevingen (zie Pendergast 1999 of Lazear en Oyer 2009 voor een overzicht).

Dat financiële beloningen een belangrijke prikkel kunnen geven zal haast niemand verbazen. Toch is het niet noodzakelijkerwijs de enige, noch de meest efficiënte manier om medewerkers te motiveren. Intrinsiek waardevol werk (Delfgaauw 2007), een sociale werksfeer of inspirerend management zijn hooggewaardeerde eigenschappen (Dur 2009). Bovendien is het essentieel om het formele beloningsbeleid en deze niet-financiële drijfveren niet in een vacuüm te bestuderen, aangezien er belangrijke interactie-effecten kunnen optreden. In dit proefschrift wordt bestudeerd hoe finan-

[1] Moral hazard is een begrip dat wordt gebruikt voor een verandering gedrag wanneer iemand de consequenties van zijn gedrag niet volledig ondervindt. In de werkplek kan dit bijvoorbeeld betekenen dat werknemers met een vast loon eerder de kantjes er vanaf proberen te lopen dan kleine zelfstandigen.

ciële prikkels invloed hebben op, en beïnvloed worden door de sociale relaties tussen collega's. Door de economische theorie op deze wijze te verrijken probeer ik meer inzicht te geven in het gedrag van organisaties en van mensen binnen de organisatie. Verder geeft dit proefschrift mogelijk bruikbare inzichten voor organisaties in hun zoektocht naar een optimaal personeelsbeleid.

Dit proefschrift presenteert twee theoretische hoofdstukken en twee hoofdstukken met de resultaten van veldexperimenten. Ik introduceer een theoretisch onderzoek naar de interactie-effecten tussen financiële prikkels en de sociale relaties tussen collega's. Hiervoor maak ik gebruik van zogeheten 'principaal-multi-agent' modellen. Deze bevatten de strategische interactie tussen werkgever en werknemer en staan formalisatie van de sociale relaties toe. De voorspellingen over mogelijke interactie-effecten tussen financiële prikkels en relaties tussen collega's zijn moeilijk om empirisch aan te tonen, omdat het problematisch is om de causaliteit vast te stellen met natuurlijke data. Zonder experiment kunnen we bijvoorbeeld geen conclusie trekken wanneer we een relatie zien tussen de kwaliteit van de relaties tussen collega's en het gebruik van teambonussen. Een bedrijf zal mogelijk enkel teambonussen introduceren wanneer de relaties goed zijn, maar de teambonus kan er ook voor zorgen dat de sfeer beter is. Veldexperimenten hebben geen problemen met de causaliteit, doordat ze gebruik maken van randomizatie. Ik beschrijf de resultaten van twee veldexperimenten over de effecten van teambonussen.

Onderstaand motiveer ik het belang van relaties tussen collega's en geef ik een korte toelichting over de voordelen en beperkingen van veldexperimenten. Daarna sluit ik af met een overzicht van de verschillende hoofdstukken uit dit proefschrift.

Relaties tussen collega's

De meeste organisaties ondernemen acties om de sociale interactie onder collega's te bevorderen. Voorbeelden hiervan variëren van de inrichting van een koffiekamer tot de organisatie van een gezamenlijke vakantie. Twee mogelijke redenen om in sociale relaties op de werkvloer te investeren zijn dat de productiviteit hoger kan worden en dat werknemers bereid zijn te werken voor een lager loon. Voor een direct effect op de productiviteit is geen doorslaggevend bewijs, maar goede relaties tussen collega's zijn wel direct positief gerelateerd aan werktevredenheid en binding met het werk en

daarnaast negatief aan stress, afwezigheid en verloop (o.a. Price and Mueller 1981, Riordan and Griffeth 1995, Nielsen et al. 2000 en Morrison 2004). Sociale interactie met collega's is een hooggewaardeerd aspect van het werk voor veel werknemers en wordt veel gemist onder gepensioneerden (Krueger and Schkade 2009 en Shacklock 2005). Dit vertaalt zich in een negatieve relatie tussen de kwaliteit van de sociale relaties binnen een organisatie en het niveau van de lonen (Borzaga and Depedri 2005).

Naast een directe investering in sociale interactie op de werkvloer hebben organisaties een invloed op de werksfeer met de financiële prikkels die ze werknemers geven. Werknemers behandelen collega's mogelijk anders wanneer hun inkomen afhangt van de inspanningen van deze collega's. Bestaande literatuur identificeert een aantal mogelijke manieren waarop financiële prikkels de acties jegens collega's beïnvloeden; een teambonus kan leiden tot meer behulpzaamheid (Fitzroy and Kraft 1986) of tot groepsdruk (Kandel and Lazear 1992) en een promotietoernooi kan leiden tot sabotage (Lazear 1989). Dit proefschrift draagt aan deze literatuur bij met een theoretisch model in hoofdstuk 2 en een veldexperiment in hoofdstuk 4.

Investeringen in de sociale interactie op de werkvloer zijn niet geheel zonder risico. Enige risico's zijn roddel, verwaarlozing van productieve taken en verstoring van het beloningsbeleid. Een voorbeeld van dit laatste punt is dat medewerkers het effect van hun inspanning op het inkomen van een bevriende collega in acht nemen. Bandiera et al. (2005) liet een productiviteitstijging zien voor een bedrijf dat relatieve beloningsprikkels verving door een individuele prikkel, deze stijging gold met name voor werknemers die met meer vrienden samenwerkten. Dit proefschrift levert een bijdrage op dit vlak in hoofdstuk 3 met een theoretische analyse van een mogelijk verstorend effect van relaties tussen collega's in 'peer evaluation', ofwel collega's die elkaar beoordelen.

Veldexperimenten

Volgens het empirisme komt kennis voort uit ervaring en observatie. Deze opvatting staat centraal in de beoefening van wetenschap, waar theorieën de empirische weerlegging moeten doorstaan. Simpele observaties zijn echter niet afdoende om een hypothese te falsificeren, aangezien men ook de 'counterfactual' moet waarnemen

voor doorslaggevend bewijs. Om bijvoorbeeld te kunnen concluderen dat bepaald beleid een positieve uitwerking heeft op de productiviteit is het niet voldoende om een stijging van de productiviteit te observeren. Deze stijging kan immers zijn veroorzaakt door factoren die niet aan het beleid gerelateerd zijn. Idealiter zou men tegelijkertijd dezelfde werknemers willen observeren in het scenario zonder het beleid, ofwel de 'counterfactual'. Bij experimenten worden 'counterfactuals' gecreëerd door middel van een willekeurige toewijzing aan een beleid; medewerkers die niet worden toegewezen vormen de controlegroep en dienen als de 'counterfactual'. Zonder experiment, ofwel met natuurlijk voorkomende observaties, is het de taak van de empirische econoom om met behulp van een identificatiemethode 'counterfactuals' te genereren; bijv. met behulp van instrumentele variabelen of matching. Veldexperimenten zijn de meeste recente toevoeging aan het gereedschap van de empirische econoom. Dit stuk over veldexperimenten pretendeert niet alomvattend te zijn, voor een uitgebreidere beschouwing van veldexperimenten in arbeidseconomie kan de lezer List en Rasul (2010) erop naslaan.

Volgens John List, de voorvechter van het veldexperiment in economie, slaan veldexperimenten een brug tussen experimenten in het laboratorium en natuurlijk voorkomende data. De interne validiteit, ofwel de geloofwaardigheid, van studies die natuurlijke data gebruiken is zo goed als de identificatiemethode die gebruikt is. Lab experimenten staan het toe om direct conclusies te trekken over causale verbanden, doordat deze gebruik maken van randomisatie. Maar lab experimenten worden ervan beticht dat ze niet representatief zijn (ofwel een lage externe validiteit hebben), bijvoorbeeld door kleine belangen, een kunstmatige omgeving, of de aanwezigheid van iemand die het experiment afneemt. Een natuurlijk veldexperiment combineert de voordelen van randomisatie en realiteit.[2]

Uiteraard zijn veldexperimenten geen wondermiddel. Wanneer de stap van het lab naar het veld wordt gemaakt gaat dat ten koste van de algehele controle over de omgeving. Daarnaast blijven er vraagtekens bestaan bij de externe validiteit, aangezien de resultaten specifiek kunnen zijn voor lokale omstandigheden. Zelfs

[2]Een natuurlijk experiment wordt gedefinieerd als een experiment waarbij de deelnemers in hun natuurlijke omgeving acteren en er niet van op de hoogte zijn dat er een experiment gaande is (Harrison en List 2004).

wanneer de resultaten van een klein veldexperiment niet locatie specifiek zijn, dan generaliseren ze niet automatisch door mogelijke evenwichtseffecten. Neem bijvoorbeeld een experiment dat laat zien dat een bedrijf productievere werknemers aantrekt wanneer het prestatiebeloning invoert; dit resultaat kan onmogelijk opgaan wanneer alle bedrijven dit doen. De meeste discussies over de voordelen en beperkingen van veldexperimenten sluiten af door te zeggen dat ze bestaande methodes aanvullen.

Tenslotte is er een aantal gevaren waar wetenschappers op bedacht moeten zijn ongeacht de empirische methode. Een eerste gevaar is de 'sample selection bias', wat wil zeggen dat de bestudeerde steekproef zodanig verschilt van de populatie dat de resultaten worden beïnvloed. De resultaten van hoofdstuk 4 en 5 zijn mogelijk ook onderhevig aan de invloed van zelfselectie. De veldexperimenten vonden namelijk plaats in één van de 15 bedrijven die we hebben benaderd. Daardoor is het mogelijk dat dit bedrijf specifieke eigenschappen heeft, zoals een management dat openstaat voor experimenten, die de resultaten kunnen beinvloeden. De invloed van zelfselectie is niet enkel van toepassing op de experimentele methode, natuurlijk voorkomende data kunnen dit probleem ook ondervinden. Banerjee en Duflo (2009) geven als voorbeeld dat grote projecten politiek gevoeliger liggen om te evalueren aangezien deze vaak goed gepubliceerd worden; daardoor kan er zelfselectie ontstaan met betrekking tot welke projecten geëvalueerd mogen worden.

Een tweede gevaar is de 'publicatie bias'. Er ontstaat een vertekening wanneer de kans op publicatie afhangt van de statistische significantie van resultaten.[3] Het achterwege blijven van nulresultaten wordt ook wel het 'ladekast probleem' genoemd, waar studies zonder significante resultaten bij wijze van beeldspraak in de ladekast verdwijnen (Rosenthal 1979). Volgens Rosenthal wordt dit probleem opgelost door replicatie, aangezien er een ongeloofwaardig grote ladekast moet zijn om de resultaten te ontkrachten. Toch is er mogelijk een probleem wanneer we de keuzevrijheid van onderzoekers m.b.t. controle variabelen combineren met een door theorie geïnspireerd vermoeden. Card en Krueger (1995) laten zien dat de literatuur over het

[3]Er zijn goede redenen voor tijdschriften om de voorkeur te geven aan studies met statistisch significante resultaten, aangezien de onderzoeksmethode of data mogelijk van hogere kwaliteit zijn. Met de 'publicatie bias' wordt echter bedoeld dat de voorkeur wordt gegeven aan studies met statistisch significante resultaten boven studies met nulresultaten van vergelijkbare onderzoekskwaliteit.

effect van minimumlonen op de werkgelegenheid mogelijk lijdt onder de publicatie bias. Hun meta-analyse toont aan dat het geschatte effect tweemaal zo groot is als de standaard fout, onafhankelijk van de grootte van het effect, en dat de t-waarden niet toenemen met de wortel van het aantal observaties. In andere woorden wordt er door alle studies slechts voldaan aan de minimale eisen van statistische significantie, terwijl de statistische significantie zou moeten toenemen met het aantal observaties. Tenslotte is de publicatie bias niet beperkt tot de voorkeur voor significante resultaten. Replicatie studies hebben meer moeite om in toonaangevende bladen te publiceren, ondanks mogelijke significantie van de resultaten. Dit probleem is mogelijk ernstiger voor experimenten. Zoals Rodrik (2009) stelt dat ironisch genoeg studies met een lagere interne validiteit een grotere prikkel tot replicatie geven onder de naam van verbeterde identificatie.

Overzicht van het proefschrift

Financiële prikkels en sociale interactie

In hoofdstuk 2 bestuderen we de mogelijkheid om sociale interactie tussen collega's te bevorderen met behulp van financiële prikkels voor productieve taken. In het principaal-agent model dat we bestuderen kiezen werknemers niet alleen productieve acties, maar ook de sociale interactie met collega's. Sociale interactie is gemodelleerd als een uitruil van aandacht, waar het geven van aandacht kostbaar is. Het krijgen van aandacht is plezierig en beïnvloedt de sociale preferenties. Werknemers zijn conditioneel altruïstisch, ofwel altruïsme richting een collega neemt toe in de ontvangst van aandacht. De private kosten en externe baten zorgen ervoor dat werknemers te weinig aandacht aan elkaar geven.

Hoofdstuk 2 laat zien dat de manager de sociale interactie kan bevorderen met een teambonus of relatieve prestatiebeloning. Deze externaliteit in een werknemer zijn loon zorgt ervoor dat er de wens is om de inspanningen van een collega te beïnvloeden. De strategische reden om aandacht aan een collega te geven werkt als volgt: Het geven van aandacht leidt tot een reciproke reactie, aangezien collega's conditioneel altruïstisch zijn. De altruïstische collega internaliseert het effect van

zijn inspanning deels en past daarmee zijn inspanning in de gewenste richting aan, als reactie op de ontvangen aandacht. De manager creëert doelbewust een externaliteit van inspanning om het probleem van de aandachts-externaliteit op te lossen. Uiteindelijk herstelt de manager de prikkels voor productieve inspanning. Hij geeft zwakkere individuele prikkels in combinatie met een teambonus, terwijl relatieve prestatiebeloning gecompeseerd wordt met sterkere individuele prikkels.

De analyse in hoofdstuk 2 leidt tot andere voorspellingen dan de bestaande literatuur. Hoofdstuk 2 voorspelt dat, onder de veronderstelling van conditioneel altruïstische of reciproke werknemers, een teambonus alsook relatieve prestatiebeloning leiden tot meer sociale interactie en betere relaties tussen collega's. Als tegenpool voorspellen theorieën van groepsdruk in teams en sabotage in promotietoernooien slechtere relaties bij deze financiële prikkels (bijv. Kandel en Lazear 1992 en Lazear 1989). Theorieën van behulpzaamheid in teams of theorieën die uitgaan van een aversie tegen ongelijkheid voorspellen ook betere relaties bij een teambonus, maar enkel als er complementariteit in de productie bestaat of wanneer de teambonus leidt tot meer inkomensgelijkheid (bijv. FitzRoy en Kraft 1986 en Bartling 2011). Naast de voorspelling dat teambonussen en relatieve prestatiebeloning leiden tot betere relaties tussen collega's heeft de analyse nog een andere beleidsimplicatie: Namelijk, dat sociale relaties van belang zijn voor het effect van financiële prikkels. Een teambonus geeft mogelijk te sterke prikkels, terwijl de prikkels te zwak zijn bij relatieve prestatiebeloning wanneer de sociale setting genegeerd wordt.

Beoordelingen door collega's en sociale relaties

In hoofdstuk 3 wordt bestudeerd hoe bestaande relaties tussen collega's invloed kunnen hebben op een optimale inzet van financiële prikkels. Ik bekijk een setting waar twee collega's betere informatie hebben over de prestaties van elkaar dan dat manager heeft. In het model betekent dit dat collega's een signaal ontvangen over de inspanningen van de ander, terwijl de manager enkel de productie van het team observeert. De manager vraagt de werknemers om hun collega te beoordelen in de hoop hiermee een sterkere prikkel te geven dan met een teambonus. De teambonus is inefficiënt doordat werknemers deels willen meeliften op de inspanningen van hun collega. Ik begin de analyse vanuit ideale omstandigheden voor 'peer evaluation'. Ik

veronderstel dat het voor werknemers kostbaar is om te liegen over het ontvangen signaal. In dit geval is een bonus voor het ontvangen van een positieve beoordeling even stimulerend als een individuele financiële prikkel zou zijn. Een eerlijke beoordeling geeft werknemers een prikkel tot inspanning, aangezien ze de kans willen verhogen dat de collega een positief signaal krijgt en daardoor een positieve beoordeling geeft.

Vervolgens introduceer ik relaties tussen collega's. Deze hebben mogelijk een verstorend effect op de beoordelingen. Werknemers maken nu een afweging tussen hun kosten van liegen en het ontleende nut van de bonus voor een (on)geliefde collega. Lage kosten van liegen, sterke sociale relaties, of grote bonussen voor een positieve beoordeling leiden er toe dat vrienden elkaar positief beoordelen ongeacht het signaal dat ze krijgen, terwijl vijanden elkaar de bonus altijd misgunnen. Door de bonus te verlagen kan de manager er voor zorgen dat de beoordelingen een eerlijke weerspiegeling van het signaal blijven. Desalniettemin, biedt de manager altijd een dergelijke bonus aan in het optimale contract. De manager vult deze enkel aan met een teambonus wanneer de sociale relaties de bonus voor een positieve beoordeling erg beperken.

Het resultaat dat relaties tussen collega's, zowel goed als slecht, de bonus voor een positieve beoordeling beperken leidt tot een testbare hypothese: 'Peer evaluation' wordt minder effectief naarmate relaties tussen collega's sterker worden. Daarnaast heeft dit resultaat een directe beleidsimplicatie voor managers: Managers moeten werknemers die relatief onverschillig tegenover elkaar staan elkaars prestaties laten beoordelen. Verder draagt hoofdstuk 3 bij aan een discussie binnen de management literatuur over de vraag of 'peer evaluation' gebruikt kan worden voor het bepalen van prestatiebeloning of dat het enkel voor educatieve doeleinden gebruikt kan worden (Edwards en Ewen 1996). Dit hoofdstuk laat zien dat sociale relaties tussen collega's geen bezwaar vormen wanneer er een kleine bonus wordt gegeven. Bovendien laat het hoofdstuk zien dat er zelfs een prikkel uitgaat van beoordelingen die het signaal negeren. Dit komt doordat werknemers de kosten van liegen van een (on)geliefde collega kunnen beïnvloedden. Als voorbeeld kan gedacht worden aan vrienden die harder gaan werken omdat ze elkaar niet in de moeilijke positie willen brengen dat ze moeten liegen over elkaar prestaties.

In de hoofdstuk 4 en 5 worden de resultaten van twee veldexperimenten beschreven. Beide veldexperimenten vonden plaats in een Nederlandse winkelketen van 128 winkels. De winkels verkopen kleding en schoenen onder één merknaam. Het management was van plan om financiële prikkels te introduceren voor het verkopend personeel, die in tegenstelling tot filiaalmanagers enkel een vast loon verdienden. Beide experimenten introduceerden een teambonus voor alle medewerkers in het filiaal bij verbeterde prestaties over een tijdspanne van zes weken. Het verkopend personeel en de filiaalmanagers waren niet op de hoogte van de samenwerking met de Erasmus Universiteit, wat deze experimenten classificeert als natuurlijke veldexperimenten.

Teambonussen en acties tegenover lakse collega's

Hoofdstuk 4 behandelt de aanleidingen voor, en consequenties van anti-shirking gedrag; ofwel acties die werknemers ondernemen wanneer ze een collega minder hard zien werken dan dat hij of zou moeten. Het hoofdstuk presenteert de resultaten van een experiment, waar alle medewerkers binnen een filiaal een bonus konden verdienen door het gemiddelde aantal producten per kassabon op winkelniveau voldoende te verhogen. Onze hypothese was dat anti-shirking gedrag zowel een effect kon hebben op de werking van de teambonus als door de teambonus beïnvloed kon worden. Vandaar dat we zowel voor als na het experiment een enquête hebben afgenomen om een indicatie van anti-shirking gedrag te krijgen.

De teambonus heeft niet tot extra verkopen tijdens het experiment geleid. Wel laat het hoofdstuk zien dat de teambonus een negatief effect heeft gehad op de bereidheid om anti-shirking gedrag te ondernemen door middel van een analyse van de verschillen tussen 'controlegroep filialen' en 'teambonus filialen' in de verschillen tussen beide enquêtes. We menen dat deze daling in de bereidheid voortkomt uit een ontevredenheid met mislukte pogingen om collega's harder te laten werken door druk uit te oefenen. Als aanwijzingen voor deze uitleg vinden we een significant effect van de teambonus op de beantwoording van twee stellingen; in 'teambonus filialen' werd vaker "de collega nam het mij kwalijk" en minder vaak "de collega in kwestie verbeterde zijn gedrag" geantwoord. Daarnaast vinden we dat de relaties tussen collega's hebben geleden onder de teambonus. We ontkrachten de alternatieve verklaring dat er sprake zou zijn van een algemene ontevredenheid in het beantwoorden van de

tweede enquête voor 'teambonus filialen', doordat we laten zien dat de relaties met de manager en werktevredenheid niet zijn afgenomen door de teambonus.

Onze analyse levert een vooruitgang op Freeman et al. (2010) door gebruik te maken van onze experimentele opzet en doordat we anti-shirking gedrag kunnen relateren aan een harde prestatiemaatstaf. In eerdere studies is laten zien dat het goed voor de productiviteit kan zijn wanneer werknemers door hun collega's worden geobserveerd (Mas en Moretti 2009 en Falk en Ichino 2006): Om goede prestaties te leveren is het belangrijk een sterke anti-shirking cultuur te creëren en te behouden. Een manier waarmee organisaties mogelijk een invloed hebben op anti-shirking gedrag is door aandacht van het management, aangezien hoofdstuk 4 laat zien dat de relaties tussen een manager en medewerkers positief gerelateerd zijn aan anti-shirking gedrag. Onze observatie dat een teambonus een negatieve uitwerking kan hebben op de anti-shirking cultuur is tegenspraak met Freeman et al. (2010).

Competitie en verschillen tussen mannen en vrouwen

Het veldexperiment in hoofdstuk 5 introduceert verkoopwedstrijden tussen groepjes van vijf filialen. In de eerste experimentele manipulatie is er een teambonus voor het winnende filiaal en de tweede plek. Terwijl er bij de tweede manipulatie slechts om de eer gestreden wordt. Filialen in beide type verkoopwedstrijden behaalden significant hogere verkoopresultaten in vergelijking met de controle groep. We vinden geen verschil tussen beide type verkoopwedstrijden. Onze bevindingen sluiten aan bij recente literatuur. Azmat en Iriberri (2010) en Blanes i Vidal en Nossol (2009) laten zien dat mensen harder werken wanneer ze feedback krijgen over relatieve prestaties en Kosfeld en Neckermann (2011) toont aan dat een competitie met enkel een symbolische prijs ook stimulerend werkt.

Daarnaast draagt hoofdstuk 5 bij aan de literatuur over verschillen tussen mannen en vrouwen in competitie. Gneezy et al. (2003) liet zien dat vrouwen minder geprikkeld worden door competitie dan mannen, waarna verschillende lab experimenten deze bevinding repliceerden. Dit verschil zou een ondervertegenwoordiging van vrouwen aan de top mede kunnen verklaren, omdat er verschillende promotietoernooien gewonnen moeten worden voor topposities. In tegenstelling tot Gneezy et al. vinden wij dat het geslacht van de filiaalmanager of het percentage vrouwen in

de winkel niet uitmaakt voor het effect van de competitie. Hiermee presenteren we het tweede veldexperiment dat geen verschillen tussen mannen en vrouwen vindt in de stimulerende werking van competitie (Lavy 2008 laat dit zien voor leraren). Dit roept de vraag op of de bevindingen van het lab naar de werkplek generaliseren.

Tenslotte vinden we dat het positieve effect van competitie afhangt van een interactie tussen het geslacht van de filiaalmanager en het percentage vrouwen in de winkel. Het stimulerende effect is enkel aanwezig wanneer er een voldoende percentage medewerkers het zelfde geslacht heeft als de filiaalmanager. Dit resultaat is interessant omdat in de competitie voor topposities, prestaties niet enkel afhangen van de eigen inspanningen van een manager, maar ook van de prestaties van het team dat hij of zij leidt. Wanneer vrouwen ondervertegenwoordigd zijn in de arbeidsmarkt dan is het volgens onze bevindingen moeilijker voor vrouwelijke managers om de top te bereiken.

Bibliography

Antonioni, David (1996), Designing an Effective 360-Degree Appraisal Feedback Process, *Organizational Dynamics*, 25(2): 24 – 38.

Antonioni, David and Heejoon Park (2001), The relationship between rater affect and three sources of 360-degree feedback ratings, *Journal of Management*, Vol. 27(4): 479-495.

Arellano, Manuel (1987), Computing Robust Standard Errors for Within-groups Estimators, *Oxford Bulletin of Economics and Statistics*, 49(4): 431 – 434.

Auriol, Emmanuelle, and Régis Renault (2008), Status and incentives, *RAND Journal of Economics*, 39(1): 305 – 326.

Azmat, Ghazala, and Nagore Iriberri (2010), The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students, *Journal of Public Economics*, 94(7-8): 435 – 452.

Bamberger, Peter A., Ido Erev, Michal Kimmel, and Tali Oref-Chen (2005),.Peer Assessment, Individual Performance, and Contribution to Group Processes: The Impact of Rater Anonymity, *Group & Organization Management*, Vol 30(4): 344-377.

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.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2009), Team Incentives: Evidence from a Field Experiment, Mimeo, University College London.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2010), Social Incentives in the Workplace, *Review of Economic Studies*, 77(2): 417-458.

Banerjee, Abhijit V. and Esther Duflo (2009), The Experimental Approach to Development Economics, *Annual Review of Economics*, 1: 151 – 178.

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

Barron, John M. and Manish M. Gupte (2009), Limits to Peer Appraisals: Peer Pressure and Team Cohesion, mimeo, Purdue University.

Bartling, Björn (2011), Relative Performance or Team Evaluation? Optimal Contracts for Other-Regarding Agents, *Journal of Economic Behavior and Organization*, forthcoming.

Baumeister Roy F. (1988), Should we stop studying sex differences altogether?, *American psychologist*, 43(12): 1092 – 1095.

Baumeister, Roy F., and Mark R. Leary (1995), The need to belong: desire for interpersonal attachments as a fundamental human motivation, *Psychological Bulletin*, 117(3): 497 – 529.

Berman, Evan M., Jonathan P. West, and Maurice N. Richter, Jr. (2002), Workplace Relations: Friendship Patterns and Consequences (According to Managers), *Public Administration Review*, 62(2): 217 – 230.

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

Bertrand, Marianne, and Kevin F. Hallock (2001), The Gender Gap in Top Corporate Jobs, *Industrial and Labor Relations Review*, 55(1): 3 – 21.

Besley, Timothy, and Maitreesh Ghatak (2008), Status Incentives, *American Economic Review*, vol. 98(2): 206 – 211.

Blanes i Vidal, Jordi, and Mareike Nossol (2009), Tournaments without Prizes: Evidence from Personnel Records, Mimeo, London School of Economics.

Bohl, Don L. (1996), Minisurvey: 360-degree appraisals yield superior results, survey shows, *Compensation and Benefits Review*, 28(5): 16 – 19.

Bolle, Friedel and Alexander Kritikos (2006), Reciprocity, Altruism, Solidarity: A Dynamic Model, *Theory and Decision*, 60(4): 371 – 394.

Booth, Alison L., and Patrick J. Nolen (2009), Choosing to Compete: How Different are Girls and Boys?, IZA Discussion Paper No. 4027.

Borgatti, Stephen P. and Pacey C. Foster (2003). The Network Paradigm in Organizational Research: A Review and Typology, *Journal of Management*, 29(6): 991 – 1013.

Borzaga, Carlo and Sara Depedri (2005), Interpersonal Relations and Job Satisfaction: Some Empirical Results, in: Benedetto Gui and Robert Sugden (Eds.), *Economics and Social Interaction; Accounting for Interpersonal Relations*, Cambridge: Cambridge University Press.

Burks, Stephen, Jeffrey Carpenter, and Lorenz Goette (2009), Performance Pay and the Erosion of Worker Cooperation: Field Experimental Evidence, *Journal of Economic Behavior and Organization*, 70(3): 458 – 469.

Card, David and Alan B. Krueger (1995), Time-Series Minimum-Wage Studies: A Meta-analysis, *American Economic Review*, 85(2): 238 – 243.

Carmichael, Lorne H. (1988), Incentives in Academics: Why is There Tenure?, *Journal of Political Economy*, 96(3): 453 – 472.

Carpenter, Jeffrey, Peter H. Matthews, and John Schirm (2010), Tournaments and Office Politics: Evidence from a Real Effort Experiment, *American Economics Review*, 100(1): 504 – 517.

Carpenter, Jeffrey and Erika Seki (2011), Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay, *Economic Inquiry*, Vol. 49(2): 612-630.

Casas-Arce, Pablo, and F. Asis Martinez-Jerez (2009), Relative Performance Compensation, Contests, and Dynamic Incentives, *Management Science*, 55(8): 1306 – 1320.

CIPD (2010), *Annual Survey Report 2010: Reward Management*, London.

Coates, Dennis E. (1998), Don't Tie 360 Feedback to Pay, *Training*, 35(9): 68 – 78.

Cohen, Don and Laurence Prusak (2001), *In Good Company: How Social Capital Makes Organizations Work*, Harvard Business School Press.

Cohn, Alain (2010), *Field Experiments in Labor Economics*, Ph.D thesis, University of Zurich.

Corneo, Giacomo and Rafael Rob (2003), Working in Public and Private Firms, *Journal of Public Economics*, 87(7-8): 1335 – 1352.

Cox, James C., Daniel Friedman, and Steven Gjerstad (2007), A Tractable Model of Reciprocity and Fairness, *Games and Economic Behavior*, 59(1): 17 – 45.

Cox, James C., Daniel Friendman and Vjollca Sadiraj (2008), Revealed Altruism, *Econometrica*, 76(1): 31 – 69.

Croson, Rachel, and Uri Gneezy (2008), Gender Differences in Preferences, *Journal of Economic Literature*, 47(2): 1 – 27.

Dargnies, Marie-Pierre (2009), Does Team Competition Eliminate the Gender Gap in Entry in Competitive Environments?, Centre d'Economie de la Sorbonne Working Paper 2009.06.

Datta Gupta, Nabanita, Anders Poulsen, and Marie-Claire Villeval (2011), Male and Female Competitive Behavior: Experimental Evidence, *Economic Inquiry*, forthcoming.

Delfgaauw, Josse (2007), *Wonderful and Woeful Work*, Ph.D thesis, Tinbergen Institute, Erasmus University Rotterdam.

Demougin, Dominique and Claude Fluet (2006), Output and Wages with Inequality Averse Agents, *Canadian Journal of Economics*, 39(2): 399 – 413.

Dohmen, Thomas, and Armin Falk (2011), Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender, *American Economic Review,* Vol 101 (2): 556-590.

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.

Ducharme, Lori J. and Jack K. Martin (2000), Unrewarding Work, Coworker Support, and Job Satisfaction, *Work and Occupations*, 27(2): 223 – 243.

Duchon, Dennis, Stephen G. Green, and Thomas D. Taber (1986), Vertical Dyad Linkage: A Longitudinal Assessment of Antecedents, Measures, and Consequences, *Journal of Applied Psychology*, 71(1): 56 – 60.

Dur, Robert, Arjan Non, and Hein Roelfsema (2010), Reciprocity and Incentive Pay in the Workplace, *Journal of Economic Psychology*, 31(4): 676 – 686.

Eagly, Alice (1995), The Science and Politics of Comparing Women and Men, *American Psychologist*, 50(3): 145 – 158.

Edwards, Mark R. and Ann J. Ewen (1996), How to Manage Performance and Pay With 360-Degree Feedback, *Compensation and Benefits Review*, 28(3): 41 – 46.

Englmaier, Florian and Achim Wambach (2010), Optimal Incentive Contracts under Inequity Aversion, *Games and Economic Behavior*, 69(2): 312 – 328.

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

Falk, Armin and James J. Heckman (2009), Lab Experiments Are a Major Source of Knowledge in the Social Sciences, *Science*, 326: 535 – 538.

Falk, Armin and Andrea Ichino (2006), Clean Evidence of Peer Effects, *Journal of Labor Economics*, 24(1): 39 – 57.

Fischbacher, Urs and Franziska Heusi (2008), Lies in Disguise. An experimental study on cheating, TWI Working Paper 40.

Fischer, Paul and John S. Hughes (1997), Mutual Monitoring and Best Agency Contracts, *Journal of Institutional and Theoretical Economics*, 153(2): 334 – 355.

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

Fleenor, John W. and Stéphane Brutus (2001), Multisource Feedback for Personnel Decisions, in: David Bracken, Carol W. Timmreck, Allan H. Church (Eds.), *The Handbook of Multisource Feedback*, San Francisco, CA: Jossey-Bass.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List (2010), Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Gender Differences in Job-Entry Decision, NBER Working Paper Number 16546.

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

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.

Frey, Bruno S., and Susanne Neckermann (2008), Awards - A View from Psychological Economics, *Journal of Psychology*, 216(4): 198 – 208.

Gandhi, Mohandas K. 1929, Our Helplessness, *Young India*, March 21.

Garicano, Luis and Ignacio Palacios-Huerta (2005), Sabotage in Tournaments: Making the Beautiful Game a Bit Less Beautiful, CEPR Discussion Paper No. 5231.

Gneezy, Uri (2005), Deception: The Role of Consequences, *American Economic Review*, 95(1): 384 – 394.

Gneezy, Uri, Ernan Haruvy, and Hadas Yafe (2004), The Inefficiency of Splitting the Bill, *Economic Journal*, 114: 265-280.

Gneezy, Uri, Kenneth L. Leonard, and John A. List (2009), Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society, *Econometrica*, 77(5): 1637 – 1664.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003), Performance in Competitive Environments: Gender Differences, *Quarterly Journal of Economics*, 118(3): 1049 – 1074.

Gneezy, Uri, and Aldo Rustichini (2004), Gender and Competition at a Young Age, *American Economic Review*, 94(2): 377 – 381.

Goel, Anand M. and Anjan V. Thakor (2006), Optimal Contracts when Agents Envy Each Other, Mimeo, DePaul University.

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

Grund, Christian and Dirk Sliwka (2005), Envy and Compassion in Tournaments, *Journal of Economics and Management Strategy*, 14(1): 187 – 207.

Gupte, Manish M. (2009), Developmental Peer Appraisals, mimeo, Purdue University.

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.

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

Hays, R B. (1988), Friendship, in: S.W. Duck (Ed.), *Handbook of Personal Relationships*, London: Wiley.

Heywood, John S., Uwe Jirjahn, and Georgi Tsertsvadze (2005), Getting along with Colleagues – Does Profit Sharing Help or Hurt?, *Kyklos*, 58(4): 557 – 573.

Heywood, John S. and Xiangdong Wei (2006), Performance Pay and Job Satisfaction, *Journal of Industrial Relations*, 48(4): 523 – 540.

Hodson, Randy (1997), Group Relations at Work, *Work and Occupations*, 24(4): 426 – 452.

Holmstrom, Bengt (1979), Moral hazard ands observability, *Bell Journal of Economics*, 10(1): 74 – 91.

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

Homans, George C. (1950), *The Human Group*, New York: Harcourt, Brace, and World.

Hurkens, Sjaak and Navin Kartik (2009), Would I Lie to You? On Social Preferences and Lying Aversion, *Experimental Economics*, 12(2): 180 – 192.

IPU (2008), *Women in Politics: 2008*, poster published by Inter-Parliamentary Union and United Nations Division for the Advancement of Women.

Itoh, Hideshi (1991a), Incentives to Help in Multi-agent Situations, *Econometrica*, 59(3): 611 – 636.

Itoh, Hideshi (1991b), Social Relations and Incentive Contracts, *Kyoto University Economic Review*, 41(1): 35 – 55.

Itoh, Hideshi (2004), Moral hazard and other-regarding preferences, *Japanese Economic Review*, 55(1): 18 – 45.

Ivanova-Stenzel, Radosveta and Dorothea Kuebler, 2005, Courtesy and Idleness: Gender Differences in Team Work and Team Competition, IZA Working Paper No. 1768.

Jurajda, Stepan, and Daniel Münich (2008), Gender Gap in Admission Performance under Competitive Pressure, CERGE-EI Working Paper no. 371.

Kandel, Eugene and Edward P. Lazear (1992), Peer pressure and partnerships, *Journal of Political Economy*, 100(4): 801 – 817.

Kartik, Navin (2009), Strategic communication with lying costs, *Review of Economic Studies*, 76(4): 1359 – 1395

Kosfeld, Michael and Susanne Neckermann (2011), Getting More Work for Nothing? Symbolic Awards and Worker Performance, *American Economic Journal: Microeconomics*, forthcoming.

Kosfeld, Michael and Ferdinand von Siemens (2011), Competition, Cooperation, and Corporate Culture, *RAND Journal of Economics,* Vol. 42(1): 23-43.

Kosfeld, Michael and Ferdinand von Siemens (2009), Worker Self-Selection and the Profits from Cooperation, *Journal of the European Economic Association*, 7(2-3): 573 – 582.

Kozlowski, Steve W. J., Georgia T. Chao, and Robert F. Morrison (1998), Games raters play: Politics, strategies, and impression management in performance appraisal, in: James W. Smither (Ed.), *Performance appraisal: State of the art in practice*, San Francisco: Jossey-Bass.

Krueger, Alan B. and David Schkade (2009), Sorting in the Labor Market: Do Gregarious Workers Flock to Interactive Jobs?, *Journal of Human Resources*, 43(4): 861 – 85.

Lavy, Victor (2008), Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-Based Pay Tournaments among Teachers, NBER Working Paper 14338.

Lazear, Edward P. (1989), Pay Equality and Industrial Politics, *Journal of Political Economy*, 97(3): 561 – 580.

Lazear, Edward P. and Paul Oyer (2009), Personnel Economics, in: Robert Gibbons and D. John Roberts (Eds.), *Handbook of Organizational Economics*, Princeton University Press, forthcoming.

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

Ledford, Gerald E., Edward E. Lawler III, and Susan A. Mohrman (1995), Reward Innovations in Fortune 1000 Companies, *Compensation & Benefits Review*, 27(4): 76-80.

Levitt, Steven D. and John A. List (2007), What do Laboratory Experiments Measuring Social Preferences tell us about the Real World, *Journal of Economic Perspectives*, 21(2): 153 – 174.

List, John A. (2006), Field Experiments: A Bridge between Lab and Naturally Occurring Data, *Advances in Economic Analysis & Policy*, 6(2): Article 8.

List, John A. and Imran Rasul (2010), Field Experiments in Labour Economics, in: Orley C. Ashenfelter and David Card (Eds.), *Handbook of Labor Economics, Vol. 4*, Elsevier.

Love, Kevin G. (1981), Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction, *Journal of Applied Psychology*, Vol. 66(4): 451-457.

Ma, Ching-To (1988), Unique Implementation of Incentive Contracts with Many Agents, *Review of Economic Studies*, 55(4): 555 – 571.

Mai-Dalton, Renate R., and Jeremiah J. Sullivan (1981), The Effects of Manager's Sex on the Assignment to a Challenging or a Dull Task and Reasons for the Choice, *Academy of Management Journal*, 24(3): 603 – 612.

March, James G., and Herbert A. Simon (1958), *Organizations*, New York: John Wiley.

Marx, Leslie M. and Francesco Squintani (2009), Individual Accountability in Teams with Adverse Selection, *Journal of Economic Behavior & Organization*, 72(1): 260 – 273.

Mas, Alexandre and Enrico Moretti (2007), Peers at Work, *American Economic Review*, 99(1): 112 – 145.

McNeilly, Kevin M., and Frederick A. Russ (2000), Does Relational Demography Matter in a Personal Selling Context?, *Journal of Personal Selling and Sales Management*, 20(2): 279 – 288.

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

Moldovanu, Benny, Aner Sela, and Xianwen Shi (2007), Contests for Status, *Journal of Political Economy*, 115(2): 338 – 363.

Morrison, Rachel (2004), Informal Relationships in the Workplace: Associations with Job Satisfaction, Organisational Commitment and Turnover Intentions, *New Zealand Journal of Psychology*, 33(3): 114 – 128.

Mossholder, Kevin W., Randall P. Settoon, and Stephanie C. Henagan (2005), A Relational Perspective on Turnover: Examining Structural, Attitudinal, and Behavioral Predictors, *Academy of Management Journal*, 48(4): 607 – 618.

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

Niederle, Muriel, and Lise Vesterlund (2007), Do Women Shy Away from Competition? Do Men Compete Too Much?, *Quarterly Journal of Economics*, 122(3): 1067 – 1101.

Nielsen, Ivy K., Steve M. Jex, and Gary A. Adams (2000), Development and Validation of Scores on a Two-Dimensional Workplace Friendship Scale, *Educational and Psychological Measurement*, 60(4): 628 – 643.

Örs, Evren, Frédéric Palomino, and Eloïc Peyrache (2008), Performance Gender-Gap: Does Competition Matter?, CEPR Discussion Paper No. 6891.

Paserman, Daniele M. (2007), Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players, mimeo, Hebrew University.

Peiperl, Maury A. (1999), Conditions for the success of peer evaluations, *International Journal of Human Resource Management*, 10(3): 429 – 458.

Pelled, Lisa Hope, and Katherine R. Xin (2000), Relational Demography and Relationship Quality in Two Cultures, *Organization Studies*, 21(6): 1077 – 1094.

Polachek, Solomom W. (1981), Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure, *Review of Economics and Statistics*, 63(1): 60 – 69.

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

Price, James L. and Charles W. Mueller (1981), A Causal Model of Turnover for Nurses, *Academy of Management Journal*, 24(3): 543 – 565.

Reagans, Ray (2005), Preferences, Identity, and Competition: Predicting Tie Strength from Demographic Data, *Management Science*, 51(9): 1374 – 1383.

Reiley, David and John A. List (2008), Field Experiments, in: Steven N. Durlauf and Lawrence E. Blume (Eds), *The New Palgrave Dictionary of Economics*, Palgrave Macmillan Publishing.

Rey-Biel, Pedro (2007), Inequity Aversion and Team Incentives, *Scandinavian Journal of Economics*, 108(2): 297 – 320.

Riordan, Christine M. and Rodger W. Griffeth (1995), The opportunity for friendship in the workplace: An underexplored construct, *Journal of Business and Psychology*, 10(2): 141 – 154.

Rob, Rafael and Peter Zemsky (2002), Social Capital, Corporate Culture, and Incentive Intensity, *RAND Journal of Economics*, 33(2): 243 – 257.

Rodrik, Dani (2009), The New Development Economics: We Shall Experiment, but How Shall We Learn? in: Jessica Cohen and William Easterly (Eds.), *What Works in Development? Thinking Big vs. Thinking Small*, Brookings Institution Press, Washington, DC.

Roethlisberger, Fritz .J. and William J. Dickson (1939), *Management and the Worker*; Harvard University Press, Cambridge, MA.

Rohlen, Thomas P. (1975), The Company Work Group, in: Ezra F. Vogel (Ed.), *Modern Japanese Organization and Decision-making*, University of California Press.

Rosen, Sherwin (1986), The Theory of Equalizing Differences. In: O. Ashenfelter and R. Layard (Eds.) *Handbook of Labor Economics, Vol. 1*, pp. 641 – 692, North Holland.

Rosenthal, Robert (1979), The "File Drawer Problem" and Tolerance for Null Results, *Psychological Bulletin*, 86 (3): 638 – 641.

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

Rotemberg, Julio J. (2006), Altruism, Reciprocity and Cooperation in the Workplace, in: S. Kolm and Jean Mercier Ythier (Eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism, Vol. 2*, pp. 1371 – 1407, Amsterdam: North-Holland.

Shacklock, Kate H. (2005), *Shall I stay?: The meaning of working to older workers in an organisational setting*, Ph.D thesis, Griffith Business School, Griffith University.

Snizek, William E., and Cecily C. Neil (1992), Job Characteristics, Gender Stereotypes and Perceived Gender Discrimination in the Workplace, *Organization Studies*, 13(3): 403 – 427.

Sonnentag, Sabine (1998), Identifying High Performers: Do Peer Nominations Suffer from a Likeability Bias?, *European Journal of Work and Organizational Psychology*, Vol. 7(4): 501-515.

Towry, Kristy L. (2003), Control in a Teamwork Environment: The Impact of Social Ties on the Effectiveness of Mutual Monitoring Contracts, *Accounting Review*, 78(4): 1069 – 1095.

Tsui, Anne S. and Bruce Barry (2009), Interpersonal Affect and Rating Errors, *Academy of Management Journal*, 29(3), 586-599.

Tsui, Anne S., and Charles A. O'Reilly III (1989), Beyond Simple Demographic Effects: The Importance of Relational Demography in Superior-Subordinate Dyads, *Academy of Management Journal*, 32(2): 402 – 423.

Van Dijk, Frans, Joep Sonnemans, and Frans van Winden (2002), Social ties in a public good experiment, *Journal of Public Economics*, 85(2): 275 – 299.

Van Dijk, Frans and Frans van Winden (1997), Dynamics of social ties and public good provision. *Journal of Public Economics*, 64(3): 323 – 341.

Van Echtelt, Patricia and Stella Hoff (2008), *Wel of niet aan het werk: Achtergronden van het onbenut arbeidspotentieel onder werkenden, werklozen en arbeidsongeschikten*, Den Haag: Social and Cultural Planning Office of the Netherlands.

Varma, Arup, and Linda K. Stroh (2001), The Impact of Same-Sex LMX Dyads on Performance Evaluations, *Human Resource Management*, 40(4): 309 – 320.

Wagner, R., Harter, J., 2006. *12: The Elements of Great Managing*, Gallup Press.

Wayne, Sandy, Robert C. Liden, and Raymond T. Sparrowe (1994), Developing Leader-Member Exchanges, *American Behavioral Scientist*, 37(5): 697 – 714.

Wirth, Linda (2001), *Breaking Through the Glass Ceiling: Women in Management*, Geneva: International Labour Office.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

452. M.I. LÓPEZ YURDA, Four Essays on Applied Micro econometrics.
453. S. MEENTS, The Influence of Sellers and the Intermediary on Buyers' Trust in C2C Electronic Marketplaces.
454. S. VUJIĆ, Econometric Studies to the Economic and Social Factors of Crime.
455. F. HEUKELOM, Kahneman and Tversky and the Making of Behavioral Economics.
456. G. BUDAI-BALKE, Operations Research Models for Scheduling Railway Infrastructure Maintenance.
457. T.R. DANIËLS, Rationalised Panics: The Consequences of Strategic Uncertainty during Financial Crises.
458. A. VAN DIJK, Essays on Finite Mixture Models.
459. C.P.B.J. VAN KLAVEREN, The Intra-household Allocation of Time.
460. O.E. JONKEREN, Adaptation to Climate Change in Inland Waterway Transport.
461. S.C. GO, Marine Insurance in the Netherlands 1600-1870, A Comparative Institutional Approach.
462. J. NIEMCZYK, Consequences and Detection of Invalid Exogeneity Conditions.
463. I. BOS, Incomplete Cartels and Antitrust Policy: Incidence and Detection
464. M. KRAWCZYK, Affect and risk in social interactions and individual decision-making.
465. T.C. LIN, Three Essays on Empirical Asset Pricing.
466. J.A. BOLHAAR, Health Insurance: Selection, Incentives and Search.
467. T. FARENHORST-YUAN, Efficient Simulation Algorithms for Optimization of Discrete Event Based on Measure Valued Differentiation.
468. M.I. OCHEA, Essays on Nonlinear Evolutionary Game Dynamics.
469. J.L.W. KIPPERSLUIS, Understanding Socioeconomic Differences in Health An Economic Approach.
470. A. AL-IBRAHIM, Dynamic Delay Management at Railways: A Semi-Markovian Decision Approach.
471. R.P. FABER, Prices and Price Setting.
472. J. HUANG, Education and Social Capital: Empirical Evidences from Micro-economic Analyses.
473. J.W. VAN DER STRAATEN, Essays on Urban Amenities and Location Choice.
474. K.M. LEE, Filtering Non Linear State Space Models: Methods and Economic Applications.
475. M.J. REINDERS, Managing Consumer Resistance to Innovations.

476. A. PARAKHONYAK, Essays on Consumer Search, Dynamic Competition and Regulation.
477. S. GUPTA, The Study of Impact of Early Life Conditions on Later Life Events: A Look Across the Individual's Life Course.
478. J. LIU, Breaking the Ice between Government and Business: From IT Enabled Control Procedure Redesign to Trusted Relationship Building.
479. D. RUSINOVA, Economic Development and Growth in Transition Countries.
480. H. WU, Essays on Top Management and corporate behavior.
481. X. LIU, Three Essays on Real Estate Finance
482. E.L.W. JONGEN, Modelling the Impact of Labour Market Policies in the Netherlands
483. M.J. SMIT, Agglomeration and Innovations: Evidence from Dutch Microdata
484. S.VAN BEKKUM, What is Wrong With Pricing Errors? Essays on Value Price Divergence
485. X. HU, Essays on Auctions
486. A.A. DUBOVIK, Economic Dances for Two (and Three)
487. A.M. LIZYAYEV, Stochastic Dominance in Portfolio Analysis and Asset Pricing
488. B. SCHWAAB, Credit Risk and State Space Methods
489. N. BASTÜRK, Essays on parameter heterogeneity and model uncertainty
490. E.GUTIÉRREZ PUIGARNAU, Labour markets, commuting and company cars
491. M.W. VORAGE, The Politics of Entry
492. A.N. HALSEMA, Essays on Resource Management: Ownership, Market Structures and Exhaustibility
493. R.E. VLAHU, Three Essays on Banking
494. N.E. VIKANDER, Essays on Teams and the Social Side of Consumption
495. E. DEMIREL, Economic Models for Inland Navigation in the Context of Climate Change
496. V.A.C. VAN DEN BERG, Congestion pricing with Heterogeneous travellers
497. E.R. DE WIT, Liquidity and Price Discovery in Real Estate Assets
498. C. LEE, Psychological Aspects of the Disposition Effect: An Experimental Investigation
499. MHA. RIDHWAN, Regional Dimensions of Monetary Policy in Indonesia
500. J. GARCÍA, The moral herd: Groups and the Evolution of Altruism and Cooperation
501. F.H. LAMP, Essays in Corporate Finance and Accounting