tinbergen institute

Jan Tichem

Altruism, Conformism, and Incentives in the Workplace

Erasmus Universiteit Rotterdam

Altruism, Conformism, and Incentives in the Workplace

ISBN: 978 90 361 0388 6

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

This book is no. 580 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.

Altruism, Conformism, and Incentives in the Workplace

Altruïsme, Conformisme en Prikkels op de Werkvloer

Proefschrift

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

De openbare verdediging zal plaatsvinden op
donderdag 15 mei 2014 om 15:30 uur
door

Johannes Tichem
geboren te Driebruggen

Promotiecommissie

Promotor:	Prof.dr. A.J. Dur
Overige leden:	Dr. J. Delfgaauw
	Prof.dr. O. Kvaløy
	Prof.dr. K.I.M. Rohde

Voorwoord

Voor u ligt mijn proefschrift, waarvan ik het bestaan enkele jaren geleden vermoedde noch ambieerde. Deze constatering roept de vraag op hoe het dan toch zo ver is gekomen. Terugkijkend zijn er twee verklarende factoren aan te wijzen. De eerste factor is een stageperiode bij het Economisch Bureau van de toenmalige Nederlandse Mededinginsautoriteit. Tijdens deze stage ontdekte ik een van de mogelijkheden die de economische wetenschap biedt om een nuttige taak te vervullen in onze wereld. Dit wakkerde bij mij de wens aan om een betere econoom te worden. De tweede verklarende factor is de uitnodiging van Robert Dur om bij hem te gaan promoveren. Het vertrouwen dat hieruit sprak gaf me de durf om de uitdaging van een promotietraject aan te gaan.

Een project met frisse moed starten is iets anders dan het naar volle tevredenheid afronden. Dat ook het laatste het geval is, is mede te danken aan de hulp van verschillende personen. Ten eerste wil ik mijn promotor Robert Dur bedanken voor zijn grote betrokkenheid bij mijn promotietraject. Robert, je deur stond altijd open en dat is voor een soms worstelende promovendus zeer prettig. Wat me ook zal bijblijven is je enorme creativiteit. In het type onderzoek dat wij verricht hebben citeert men nogal eens uit de management literatuur. Een veelgehoord inzicht uit die hoek is dat het van belang is om 'problemen te veranderen in kansen'. Volgens mij ben jij een van de weinige mensen die dat echt kan en doet. Met behulp van deze vaardigheid heb je me niet alleen door moeilijke momenten uit mijn promotietraject geholpen, maar geef je ook anderen veel opbouwende feedback. Jouw voorbeeld neem ik mee in de toekomst.

Verder wil ik graag Josse Delfgaauw en Kim Rohde bedanken voor hun deelname aan de kleine commissie. I would also like to thank Ola Kvaløy for taking part

in the doctoral committee. Daarnaast gaat mijn dank uit naar Otto Swank en Jarig van Sinderen voor hun deelname aan de grote commissie. Jarig, naast de voornoemde stage blijkt er voor ons meer samenwerking op het Economisch Bureau in het verschiet te liggen. Ik hoop en verwacht dat we een prettige en productieve tijd tegemoet gaan.

Robin, ook jou ben ik dankbaar voor de afgelopen jaren. We zijn ongeveer tegelijk begonnen aan ons promotietraject en werkten aan vergelijkbare onderwerpen. In vele gesprekken scherpte je mijn gedachten over mijn onderzoek aan, en heb ik veel geleerd jouw onderzoek. Daarnaast heb ik veel plezier beleefd aan onze talloze discussies over gerelateerde en minder gerelateerde onderwerpen. Wat ik ook als zeer waardevol heb ervaren is dat we elkaar een steuntje in de rug konden geven als het even wat minder lekker liep. Tenslotte was het bezoek aan de EEA/ESEM conferentie in Málaga een van de hoogtepunten. Ik wens je heel veel succes bij de laatste fase van je promotietraject en bij het vinden van je volgende leuke baan. Michiel en Max, ook aan jullie denk ik met plezier terug. De gezelligheid op H7-16 en de unieke kwaliteit van onze grappen zal ik zeker gaan missen. Ik wens jullie het allerbeste toe in het verdere van jullie promotie. Verder ben ik blij dat ik op de Erasmus Niels ontmoet heb. Niels, tijdens menige lunch hebben we gereflecteerd op het bestaan van een promovendus. Daarin heb ik genoten van de theologisch/wijsgerige wending die het gesprek wel eens nam.

Een aantal mensen is op meer indirecte wijze betrokken geweest bij mijn promotietraject. Voor hun inzichten, hulp en aanmoediging wil ik bedanken Oke, Heiner, Jurjen, Suzanne en Arjan.

Onmisbaar in de afgelopen jaren was mijn vrouw. Annie, jij weet als geen ander hoe veeleisend het promotietraject is geweest. Wetenschappelijk werk neemt men mee naar huis, waardoor ik nogal eens afwezig was. Een sprekend voorbeeld in dezen is dat ik soms van tafel liep om een gedachte over een prangend probleem uit te werken. Ik ben je heel dankbaar voor je geduld de afgelopen jaren. Daarnaast kon ik altijd de 'menselijke kant' van het promotietraject met je bespreken. Je nuchtere kijk hielp me om gemotiveerd en gefocust te blijven, en te genieten van de afgelopen jaren. Het is een van de redenen dat ik van je hou.

Tenslotte dank ik God voor Zijn zegeningen, waarvan hierboven een grote verscheidenheid is genoemd.

Contents

Chapter 1

Introduction

1.1 Motivation, aim, and method

To be successful, organizations need their employees to do their best. However, for two reasons employees may not do that. The first reason is that labor relationships are characterized by a divergence of interests: putting in hard work benefits the employer, whereas the costs of it are borne by the employee. It is therefore in the employee's interest to shirk. Second, it is usually impossible to force employees to work hard by means of a legal contract. This holds because in most cases employees' effort cannot be verified by a judge. These premises, which are the basic premises of principal-agent theory, imply that organizations must motivate their employees to act in the organization's interest (see e.g. Laffont and Martimort 2002).

Traditionally, economists stress the importance of monetary incentives for motivation. This is not without merit, since well-designed studies show that employees improve their performance when doing so increases their compensation (see e.g. Lazear 2000 and Shearer 2004; for overviews of the literature see Prendergast 1999 and Lazear and Oyer 2013). Nevertheless, money is not the only motivator in the workplace. One of the most compelling pieces of evidence in this regard, is that for many if not most employees, compensation is not directly linked to their performance (see e.g. MacLeod and Parent 2000). Moreover, if compensation is linked to performance, the share of performance-related compensation is typically low. Economists recognize these facts, and over the last decades they have developed a large literature

studying non-monetary motivations in the workplace. Among them are reciprocity (e.g. Akerlof 1982 and Non 2012), intrinsic motivation for specific tasks (e.g. Besley and Ghatak 2005 and Delfgaauw and Dur 2007), conformism (e.g. Sliwka 2007), altruism (e.g. Prendergast and Topel 1996), and inequity aversion (e.g. Fehr and Schmidt 1999 and Bolton and Ockenfels 2000).

The present thesis contributes to the economic literature on non-monetary motivations in the workplace. The aim is to study the importance of altruism and the preference to conform to social norms for employee motivation. Chapters 2 and 3 study the implications of altruism and spite between an employee and his employer or manager for optimal incentive contracts. Both chapters focus on the case where employee performance is subjectively evaluated. Chapter 4 studies how the preference to conform to social norms affects employee motivation, and derives implications for organizational design. The results of the present thesis are obtained by developing and analyzing principal-agent models. Because the research conducted here is theoretical in nature, its scientific value lies primarily in generating new hypotheses and interpreting observed facts. However, some of the insights also have practical implications for the workplace.

The remainder of this introduction is organized as follows. In the next subsection, I discuss the concepts of altruism and conformism and motivate their importance for the workplace. Section 1.3 provides an overview of the thesis.

1.2 Altruism and conformism

The first key behavioral assumption employed in this thesis is that people may be altruistic. The Oxford English Dictionary defines 'altruism' as a "disinterested and selfless concern for the well-being of others". The body of evidence suggesting that at least some people are altruistic is large. For example, in laboratory experiments, people are often found to cooperate in prisoners' dilemmas. More evidence from the laboratory is provided by the numerous experiments involving the dictator game. In this game, one player decides how to split a fixed sum of money between himself and one other person. In contrast to the prediction of the classic economic model,

where people are assumed to be selfish money-maximizers, usually 60% of the players give some money to their opponent. The average size of the transfer is 20% of the money at stake (Levitt and List 2007). A last example from the laboratory comes from observed behavior in public good games. In this type of game, each player chooses how much money to contribute from her endowment to a common pool. The common pool is divided among all players of the game, after its value has been multiplied by a factor greater than one. For this reason, it is optimal for the group if each player contributes his entire endowment to the pool. However, if one behaves as a selfish money-maximizer, it is optimal to contribute nothing to the pool. The reason is that each player incurs the full costs of his contribution, but since the pool is divided among n players, reaps only a fraction $1/n$ of its benefits. Nevertheless, total contributions typically lie between 40% and 60% of the group optimum (Ledyard 1995). This suggests that players take into account the well-being of others. The laboratory evidence described above not only suggest that some people are altruistic, but also that some people are negatively altruistic, or spiteful. For example, Andreoni and Miller (2002) find that 55% of their experimental subjects are willing to give up money to increase someone else's payoff, whereas 23% are willing to give up money to reduce someone else's payoff.

Whether the laboratory evidence described above constitutes a definitive proof that some people are altruistic has been heatedly debated. One reason is that empirical studies show that presumably altruistic behavior may be driven by other concerns. For example, Landry et al. (2006) conduct a field experiment on charitable giving in door-to-door fundraising campaigns, and find that donations increase in the attractiveness of the female solicitors. In a similar setting, DellaVigna et al. (2012) exogenously vary the presence of a flyer on the doorknob notifying the time that solicitors will visit. In this flyer treatment, people are significantly less likely to open the door, which is consistent with the hypothesis that people give to charities because of social pressure. However, DellaVigna et al. (2012) do not find that the share of large donations (higher than \$10) is affected by the flyer treatment. The authors take this as indirect evidence in favor of the existence of altruism.

The difficulty that lies at the heart of the debate on altruism is that its exis-

tence cannot be verified. One can observe someone's behavior, but not the motives driving it. As a result, one can never rule out that presumably altruistic behavior is actually driven by some other ulterior motive. Experimental studies on altruism have therefore tried to falsify the possibility of altruism by ruling out as much as possible other ulterior motives. This strand of literature finds that the behavior observed in dictator games and public good games can be partly explained by players' incentives to form a reputation, the warm glow of giving, reciprocity, and social pressure. However, this type of research cannot rule out the possibility of altruism. The laboratory evidence described above therefore serves as a strong indication that some people are altruistic or spiteful (see Andreoni et al. 2008 for a review of the literature).

There is also more specific evidence relating to the importance of altruism between employees and their employers or managers. Surveys among managers reveal that friendships between managers and employees occur frequently (see for instance Berman et al. 2002). Furthermore, Campbell and Kamlani (1997) find that a large majority of US compensation managers deem good quality manager-employee relations more important in determining effort than good working conditions, high wages, and monitoring.* Bad relationships between employees and managers do also occur. Moerbeek and Need (2003) report Dutch data showing that in eight percent of the jobs respondents had in their lives, they got along with their manager badly or very badly. More tentatively, Kahneman et al. (2004) report diary evidence from a US sample of employed women showing that of all regular daily activities, respondents dislike most to interact with their boss.

Altruism between employees and their employers or managers has various implications for labor relationships. For example, in case of employer-employee relationships, altruism may reduce motivation problems in effort provision. The reason is that altruistic employees enjoy to contribute to their employer's well-being. Following this logic, a theoretical literature establishes that altruistic employees are optimally provided lower incentives (e.g. Casadesus-Masanell 2004, Sliwka 2007,

*In line with this, an extensive literature in organizational psychology has established a strong positive correlation between the quality of manager-employee relationships and employee performance (see e.g. Wayne et al. 1997 and Rhoades and Eisenberger 2002).

Shchetinin 2010, and Non 2012). Another example is that good social relationships on the workfloor make jobs more valuable to employees. Theory therefore suggests that good social relationships on the job are a substitute for high compensation (e.g. Grund and Sliwka 2005, Arce 2010, and Dur and Sol 2010). Some empirical evidence for this prediction is reported by Borzaga and Depedri (2005). For the Italian non-profit sector, the authors find a negative correlation between the quality of social relations on the workfloor and wages.

The second key behavioral assumption employed in this thesis is that people have a preference to conform to social norms. An early, among economists infamous experiment documenting this phenomenon is Asch (1951). In this experiment, subjects had to perform the easy task of matching lines of corresponding length. Accuracy rates reached 98% in case subjects performed this task individually. In the treatment, subjects had to perform the task in groups of eight. The subjects were deceived in that four of the eight group members were confederates of the experimenter, who unanimously gave a wrong answer. Accuracy rates drop to less than two-thirds, which suggests that subjects want to avoid disagreement with the judgment of others.

Over the last decade, a number of well-designed empirical studies have provided evidence on the importance of conformism for the workplace. Ichino and Maggi (2000) show that employees engage less in misconduct and absent themselves less often if their co-workers are less likely to do so (also see Bradley et al. 2007). Falk and Ichino (2006) report that, if employees work in pairs rather than individually, the standard deviation of output is smaller and output is higher. Moreover, the authors find that low-productivity employees respond strongest to working in pairs. Mas and Moretti (2009) find that supermarket cashiers improve their performance if a high-productivity cashier enters their shift. As in Falk and Ichino (2006), this effect is far more pronounced for low-productivity workers than for high-productivity workers. Mas and Moretti (2009) provide a social pressure interpretation for their results, based on the finding that cashiers only improve their performance if this can be observed by the new cashier. Finally, in a controlled work environment, Bradler et al. (2013) show that the provision of public recognition to employees improves

their performance. Similar to Falk and Ichino (2006) and Mas and Moretti (2009), the authors report that their results are mainly driven by those employees who did not receive recognition, and thus learned that they performed worse than others. For this reason, Bradler et al. (2013) explain their results partly in terms of employees' preferences to conform to a group norm.

Finally, even though the assumptions of altruism and conformism are departures from the classic economic model, which assumes that people behave as selfish money-maximizers, the assumption that people behave rationally is maintained.[†] The models developed in this thesis therefore describe the preference for altruism or to conform by means of utility functions, which are assumed to be maximized by the players involved. As a result, the models developed here compare closely to classic economic models, except that they allow for other ends besides money to be pursued.

1.3 Overview of the thesis

Chapters 2 and 3 of this thesis study how altruism and spite between an employee and his employer or manager affect optimal incentive contracts. In both chapters, it is assumed that the employee's performance is subjectively evaluated. Subjective performance evaluation is widely observed in practice (see e.g. Gibbons 1998, Prendergast 1999, Gibbs et al. 2004, and Gibbs 2012). In such cases, it is not only impossible to write an explicit contract on the employee's effort, but also on his performance. The reason is that a subjective evaluation cannot be verified by a judge. Employers must therefore revert to alternative ways to incentivize their employees. In the remainder of this chapter, I will refer to employers as principals, to managers as supervisors, and to employees as agents.

In chapter 2, Robert Dur and I study the implications of altruism and spite between a principal and an agent for optimal relational incentive contracts. A rela-

[†]Andreoni and Miller (2002) experimentally test the rationality of preferences for altruism by varying players' endowments and the price at which money can be transferred to another player. The authors conclude that most of their subjects make altruistic or spiteful choices that do not violate the Generalized Axiom of Revealed Preference.

tional incentive contract motivates the agent by means of promises and/or threats based on the principal's subjective evaluation of the agent's performance. Because of the subjectivity involved, a relational contract cannot be enforced in court. To be effective, a relational contract must therefore be credible. Chapter 2 contributes to two large strands of literature. On the one hand, it adds to the literature studying relational contracts by accounting for the presence of altruism and spite between the contracting players (see e.g. Bull 1987, MacLeod and Malcomson 1989, Baker et al. 1994, and Levin 2003). On the other hand, the chapter adds to the literature studying the implications of social preferences for optimal incentive contracts (see e.g. Grund and Sliwka 2005, Dur and Glazer 2008, and Non 2012).

In the model, we allow the principal to use two kinds of incentives: a promise to pay a bonus if the agent's performance is good, and a threat of dismissal if the agent's performance is bad. Both the principal and the agent have some bargaining power, implying that they divide the total value of the labor relationship between each other. The analysis reveals that principal's and agent's altruism improves the credibility of a promise to pay a bonus, but undermines the credibility of a threat of dismissal. These results imply two counterintuitive comparative statics. First, in contrast to the existing literature (e.g. Casadesus-Masanell 2004, Sliwka 2007, Shchetinin 2010, and Non 2012), we find that more altruism may lead to higher bonuses. The reason is that higher altruism may be required to make a bonus credible in the first place. Second, even though altruism implies that the players enjoy each other's well-being to some extent, a marginal increase in altruism may make players worse off in equilibrium. The intuition is that higher altruism may render a threat of dismissal not credible. If this holds, the agent may face lower incentives to exert effort, in which case the production value generated within the labor relationship is reduced.

Whereas chapter 2 studies altruism and spite between a principal and an agent, in chapter 3 I turn to the case where a non-residual claimant supervisor may be altruistic toward an agent. I assume that the principal uses the supervisor to evaluate the agent's performance, and incentivizes the agent on the basis of the supervisor's report. The supervisor may be one of two types: selfish, or altruistic toward the

agent. In the model, supervisor's altruism may affect the agent's incentives to exert effort. The reason is as follows. Since an altruistic supervisor cares for the agent's well-being, he may find it optimal to report high performance irrespective of the agent's actual performance. However, if the agent expects this, he has no reason to exert effort. To be able to incentivize the agent, the principal must therefore make sure that the supervisor reports performance truthfully. In chapter 3, I derive the optimal incentive contracts for the supervisor and the agent that ensure truthful performance evaluations.

The agency problem described above was first studied by Prendergast and Topel (1996). The authors find that, to reduce the costs of supervisors playing favorites, firms optimally limit agents' incentives and use bureaucratic rules in pay and promotion decisions. However, they do not derive optimal contracts for the supervisor and agent that induce truthful performance evaluations from the supervisor. Giebe and Gürtler (2012) do study this issue in a one-shot game where the agent is assumed not to know the supervisor's type. In contrast to their model, I assume that the supervisor and agent interact during more periods and, as a consequence, the agent knows the supervisor's type. In addition to this, the principal can hire one agent and one supervisor from a large pool of workers, and the principal can offer contracts to a new agent and supervisor every period. Finally, I assume that the contract offered to the supervisor is uniform, that is, the principal cannot offer a menu of contracts to the supervisor.

The analysis in chapter 3 yields two novel insights. First, in contrast to the existing literature, it reveals that supervisor's altruism may provide a net incentive to report performance truthfully, rather than to bias evaluations. As in the models mentioned above, supervisor's altruism increases the benefit of biasing the performance evaluation upwards. However, in contrast to these models, I show that altruism also increases the benefits of reporting truthfully. The reason is that, because of his altruism, the altruistic supervisor enjoys a rent from working with the agent. I show that the optimal supervisor's contract ensures that the supervisor loses this rent, however, once he rates the agent's performance incorrectly. Therefore, if the supervisor cares sufficiently for the future, supervisor's altruism provides a net

incentive to report performance truthfully. Second, it may be optimal to design the supervisor's contract such that it attracts only one of the supervisor types. The intuition is that such a screening contract, conditional on being accepted, reveals the supervisor's type. I show that possessing this information enables the principal to incentivize the supervisor to report performance truthfully at lower costs.

In chapter 4, I study the behavioral assumption that agents have a preference to conform to social norms for their effort. The aim of the chapter is to analyze how hierarchical firms optimally create effort norms for their agents. There is a small but growing theoretical literature that studies conformism in the workplace. Currently, the focus is mainly on the implications of conformism for optimal incentive schemes. For example, Sliwka (2007) develops a model where the principal's choice whether or not to offer financial incentives may affect the agent's perception of the social norm. Fischer and Huddart (2008) show that the effectiveness of individual incentives is multiplied if agents have a preference to conform. The contribution of chapter 4 to the existing literature lies in its novel assumption that norms within firms are affected by the actual behavior of superiors. To study the implications of this assumption for optimal norm creation in hierarchical firms, I develop a principal-supervisor-agent model. In the model, each player exerts effort which yields valuable production for the principal's firm. The principal's effort sets the norm for the supervisor's effort, which in turn sets the norm for the agent's effort. The supervisor and the agent incur a psychological cost if their effort falls below the norm.

The key result of chapter 4 is that hierarchical firms are confronted with 'norm-erosion'. That is, effort norms decrease as one moves down the hierarchy. The mechanism driving this result is simple. When exerting effort, the supervisor takes into account the norm set by the principal, but never fully complies with it. The reason is that it is privately optimal to trade off some costs of performing below the norm against the benefit of some shirking. The implication of norm erosion is that the agent faces an inefficiently low norm. The second result is that the principal optimally exerts relatively much effort, and thereby sets a relatively high example for the supervisor. The benefit is that the agent faces a more efficient norm in equilibrium. However, this is a costly strategy for two reasons. First, the principal

incurs high effort costs. Second, the supervisor faces a high norm, implying that he incurs high effort costs and norm violation costs in equilibrium. The supervisor must be compensated for these costs through his wage. In equilibrium, hierarchical firms therefore incur three inefficiencies in norm creation: the principal's effort is inefficiently high, the supervisor's norm is inefficiently high, and the agent's norm is inefficiently low. Finally, the analysis shows that to reduce these inefficiencies, hierarchical firms optimally keep the extent of hierarchy to a minimum, promote agents with the strongest sensitivity to social norms, and distort supervisors' spans of control.

Chapter 2

Altruism and Relational Incentives in the Workplace

Joint with Robert Dur[1]

2.1 Introduction

Incentive contracts for workers often do not rely on objective performance measures only. Indeed, "thinking of *any* job in which subjective evaluation or supervisor discretion does not play some role in incentives is difficult" (Gibbs 2012, p. 15). Subjective performance evaluation sometimes affects pay. For example, 34% of employees in the industrial sector in the UK received some form of merit pay "which depended on a subjective judgement by a supervisor or manager of the individual's performance" (quoted by MacLeod and Malcomson 1998 from Millward et al. 1992, p. 388).[2] However, the use of subjective performance evaluation is not restricted to bonus pay only. Managers and employees regularly have an understanding that the employment relationship is only continued if performance is satisfactory, which is often a subjective matter. If the job is valuable to a worker, such a threat of dismissal also works as an incentive device (Shapiro and Stiglitz 1984). A prominent

[1] A slightly adapted version of this chapter is forthcoming in *Journal of Economics and Management Strategy*.

[2] See MacLeod and Parent (2000) and Gibbs et al. (2004) for similar evidence concerning subjectively determined bonuses in other sectors.

example is Henry Ford's five-dollar-day program which almost doubled wages (Raff and Summers 1987). Indirect evidence that many firms use efficiency wages as an incentive device is that bonuses are more common when the unemployment rate is low, that is, when having a job is less valuable to a worker (MacLeod and Parent 2000).

Subjective performance evaluation can overcome some well-known problems related to the use of objective performance measures, such as multi-tasking concerns, measurement costs, and lack of flexibility. When such problems are severe, managers may revert to 'relational contracts' in which employee performance is evaluated subjectively in a holistic way (Gibbons 1998, Prendergast 1999). However, a difficulty with relational contracts is that they cannot be enforced in court, but instead must be self-enforcing. Promises and threats contained in a relational contract may therefore be weak as they are constrained by their credibility. In recent decades, a rich theoretical literature has developed studying the optimal design and use of relational contracts (see among others Bull 1987, MacLeod and Malcomson 1989 and 1998, Baker et al. 1994, and Levin 2003).

The present chapter contributes to this literature by studying how altruism between managers and employees affects the optimal design of relational contracts. To this end we develop a simple dynamic principal-agent model where both players have some bargaining power. A moral hazard problem exists because both the agent's effort and performance are non-verifiable. For this reason, contracts that condition on effort or performance are not enforceable in court, and must therefore be self-enforcing. The contract may contain two types of incentives for the agent to work hard: a promise to pay a bonus for good performance as in e.g. Baker et al. (1994), and a high wage combined with a threat of dismissal following bad performance (efficiency wages) as in Shapiro and Stiglitz (1984). Our key innovation is that we allow both players to have feelings of altruism and/or spite toward each other. These feelings need not be symmetric. Our analysis yields several potentially testable hypotheses as well as a number of management implications.

So far, most studies on relational contracts have abstracted from altruism or spite between the contracting parties. This is somewhat surprising given the prevalence

of relational contracts in the workplace, as described above, and the abundance of evidence for the existence of altruism and spite between managers and employees. Surveys among managers reveal that friendships between managers and employees occur frequently (see for instance Berman et al. 2002). Furthermore, Campbell and Kamlani (1997) find that a large majority of US compensation managers deem good quality manager-employee relations more important in determining effort than good working conditions, high wages, and monitoring.[3] There is also evidence for the occurrence of bad manager-employee relationships. Moerbeek and Need (2003) report Dutch data showing that in eight percent of the jobs respondents had in their lives, they got along with their manager badly or very badly. More tentatively, Kahneman et al. (2004) report diary evidence from a US sample of employed women showing that of all regular daily activities, respondents dislike most to interact with their boss.[4]

Some of the results from the present analysis can best be compared against the benchmark of the traditional relational contracts literature, which commonly assumes selfish preferences (see the next section for a review of the related literature). In this respect, our first key result is that higher altruism (meaning that either the principal, the agent, or both are more altruistic) improves the credibility of a promise to pay a bonus, ceteris paribus. The reason is twofold. First, higher altruism makes the relational contract more valuable, which gives the principal stronger incentives to adhere to it. Second, an altruistic principal partly internalizes the benefits of the bonus to the agent, which reduces the principal's costs of honoring the contract. The second key result is that higher altruism undermines the credibility of a threat of dismissal. Intuitively, altruism makes the employment relationship more valuable, also if the worker shirks. As a consequence, the principal may find it attractive to retain a worker with whom she has good ties even if the worker shirks, implying that a threat of dismissal is not credible. Practitioners seem to be well-aware of

[3] In line with this, an extensive literature in organizational psychology has established a strong positive correlation between the quality of the manager-employee relationship and employee performance (see e.g. Wayne et al. 1997 and Rhoades and Eisenberger 2002).

[4] There is also substantive laboratory evidence indicating that a majority of people are altruistic, even to strangers, see the review by Andreoni et al. (2008). However, other people are actually spiteful. For instance, Andreoni and Miller (2002) find that 55% of their sample is altruistic while 23% is spiteful.

the impact of altruism on the credibility of incentives. In line with our first key result, it has been claimed that managers can more credibly make promises if they are "individuals whom others perceive to be [...] likeable - they're pleasant. You enjoy being around them. You may not necessarily want them for your best friend, but you wouldn't mind having a meal or drink with them.".[5] Corroborating our second key result, others have recognized that "to fire a friend is one of the hardest tests of leadership...".[6]

Further, we find that altruism and spite have the following implications for optimal relational contracts. First, in case of strong feelings of spite, players don't contract even though the economic surplus from trading is positive. This holds because feelings of spite imply a social cost from entering into or continuing an employment relationship.

Second, in case of strong feelings of altruism, no incentives are required at all to make the agent work hard. The reason is that an altruistic agent enjoys enriching the principal. Hence, the optimal contract is a flat-wage contract without a threat of dismissal that, nevertheless, results in high effort.

Third, in the presence of moderate feelings of altruism, both incentive types are credible. In that case, the use of both incentives is sometimes necessary to induce high effort. However, it can also happen that players are indifferent between using both incentives or only one of them. The reason is that either the principal's rents may be so large that she can credibly promise a bonus which is sufficiently high to induce high effort, or the job may be so valuable to the agent that a threat of dismissal provides sufficiently strong incentives, or both.

Fourth, for sufficiently high altruism, a threat of dismissal is no longer credible and so the only available incentive is a promise to pay a bonus. The optimal relational contract induces high effort through bonus pay if the discount factor is sufficiently high. If the latter condition is not satisfied, neither a threat of dismissal

[5]Tony Simons and Liz Guthridge, "How to Help Your Leaders be Credible in Incredible Times." *Integrity Dividend LLC*, August 1, 2010, accessed April 4, 2013, http://integritydividend.com/leadership/how-to-help-your-leaders-be-credible-in-incredible-times.html.

[6]Robert Bruner, "To Fire a Friend." *The Economic Times*, March 16, 2012, accessed March 19, 2013, http://articles.economictimes.indiatimes.com/2012-03-16/news/31201692_1_managers-drinking-friends-assassination-attempt.

nor a bonus is credible, but players still enter into an employment relationship. The reason is that, if a threat of dismissal is not credible, it is by definition attractive to hire the agent even if he shirks. Hence, the optimal contract has no incentives and results in low effort.

The remaining results from our analysis can best be compared with the literature that examines how social preferences affect the use of incentives in one-shot games (this strand of related literature is also discussed in the next section). Starting from the low-effort contract described above, we find the following counterintuitive comparative static results. First, higher altruism can lead to higher bonuses. This follows directly from the first key result that altruism improves the credibility of a bonus. Second, lower altruism can lead to higher effort. This holds because lower altruism may make a threat of dismissal credible, which implies that it becomes possible to induce high effort. As a result, a marginal decrease in altruism can increase players' utility in equilibrium.

Summarizing, the interaction of altruism and relational incentives yields a number of unique insights. First, whether altruism improves the credibility of relational incentives depends crucially on the kind of incentive that is being used. Second, higher altruism may lead to higher bonuses. Last, players may benefit from lower altruism when this makes a threat of dismissal credible.

The remainder of the chapter is structured as follows. Section 2.2 describes the related literature. Section 2.3 describes the model. Section 2.4 contains the main analysis and results, after which section 2.5 concludes with a discussion of management implications, testable predictions, and directions for further research.

2.2 Related literature

Most closely related to the present analysis is the contribution by Cordero Salas and Roe (2012), who study relational contracting between a selfish agent and an altruistic principal. Like us, they find that the credibility of the bonus increases in the principal's altruism. This chapter differs in that we allow not only the principal, but also the agent to have feelings of altruism or spite. We show that the optimal

bonus critically depends on the agent's altruism. Moreover, as both players have some bargaining power, the relational contract in our model may contain a threat of dismissal, which is a substitute for bonus pay. Consequently, and unlike Cordero Salas and Roe (2012), also the principal's altruism affects the optimal level of the bonus.

Another closely related paper is by Kragl and Schmid (2009), who analyze the impact of disadvantageous inequity aversion (or: envy) between two agents on the credibility of the principal's promise to pay a bonus conditional on individual performance. The authors show that the effect is ambiguous, implying that agent's envy can be beneficial to the principal (see also Kragl 2009 and 2011). This chapter differs in a number of important ways. First, our focus is on altruism and spite rather than envy. Second, unlike Kragl and Schmid (2009), the agent's feelings of altruism and spite are directed towards the principal rather than to co-workers, and the principal may be altruistic or spiteful towards the agent as well. Last, in our analysis the agent holds some bargaining power. As a result of these differences, we obtain several new insights. For instance, whereas Kragl and Schmid (2009) show that the feasibility of a relational contract may require agents to be sufficiently envious towards each other, we find that the principal and agent must be sufficiently altruistic (or not too spiteful) towards each other. Moreover, unlike Kragl and Schmid (2009), we provide a rich set of results on the interaction between the use of a threat of dismissal and bonus pay, and how this is affected by altruism.

Also related to our analysis is Brown et al. (2004) who formally derive, and confirm in the laboratory, the hypothesis that if some unobserved fraction of agents is reciprocal, long-term relational contracts will emerge that generate high effort levels, exhibit rent-sharing, and punish low effort with dismissal. The reason is that principals pay rents in order to motivate reciprocal agents which, in turn, gives selfish agents an incentive to mimic reciprocal workers (see also Brown et al. 2012). This chapter differs in that we assume that players are unconditionally altruistic or spiteful, types are observable, and we focus on one principal-agent pair. Importantly, we also allow the principal to have feelings of altruism or spite.

This chapter is also related to the literature that studies the interaction between

social preferences and incentives in static models. The theoretical literature can be divided into two groups of papers, one studying 'horizontal' social preferences and the other studying 'vertical' social preferences. The first group focuses on situations where people have social preferences towards people who are in the same hierarchical level, like in co-worker relationships.[7] The second group considers, as we do, social preferences towards people in a different hierarchical level, like in manager-employee relationships.

A seminal paper in the field is Prendergast and Topel (1996)'s model of favoritism in organizations. In their principal-supervisor-agent model, the supervisor subjectively evaluates the agent's performance and reports this to the principal. The supervisor's report is the basis for the agent's remuneration. Further, the supervisor may be altruistic or spiteful towards the agent, which leads him to distort his performance reports. In a static context, Prendergast and Topel (1996) study a rich set of issues including optimal performance pay for employees, the extent of the supervisor's authority, and the use of bureaucratic rules in pay and promotion decisions. Lee and Persson (2011) also develop a three-layer model and allow for two-sided altruism between a supervisor and agent. They show that although supervisor's altruism induces leniency, agent's altruism induces loyalty such that social relationships in the workplace may well be in the interest of the principal. Moreover, they show that when social relationships intensify, the dominant governance mode shifts from a controlling authority regime to a trusting loyalty regime.[8] We differ from these papers in two important ways. First, we assume that a residual claimant principal evaluates the agent's performance. Second, we also study the optimal use of a threat of dismissal.

[7] Models have been developed studying the effect of envy among employees on piece rates (Bartling and Von Siemens 2010a), tournament incentives (Grund and Sliwka 2005, Bartling 2011), and team incentives (Rey-Biel 2008, Bartling and Von Siemens 2010b). Other models assume horizontal inequity aversion and derive implications for optimal piece rates (Englmaier and Wambach 2010, Neilson and Stowe 2010), tournament incentives (Itoh 2004, Ederer and Patacconi 2010), and team incentives (Itoh 2004, Englmaier and Wambach 2010). Von Siemens (2011 and 2012) study screening contracts for inequity averse workers. Sol (2010) studies the optimal use of peer evaluations when co-workers are altruistic or spiteful towards each other. Last, Rotemberg (1994) and Dur and Sol (2010) study how incentive contracts can affect employees' willingness to invest in co-worker altruism.

[8] Also see chapter 3 of this thesis where, in a dynamic setting, I derive optimal contracts that induce a potentially altruistic supervisor to truthfully report employee performance.

In different contexts, Casadesus-Masanell (2004), Sliwka (2007), Shchetinin (2010), and Non (2012) have studied incentive provision to altruistic employees by (possibly) altruistic managers. A common result – that we will also obtain in our analysis – is that altruistic employees require a lower bonus as they enjoy enriching the principal. Analogously, Itoh (2004), Dur and Glazer (2008), and Englmaier and Wambach (2010) study optimal incentive contracts for employees who are envious or inequity averse (that is, care about the difference between their manager's and their own payoff).[9] As in our analysis of spiteful employees, such social preferences are a disincentive to provide effort. Pay-for-performance weakens this disincentive, as employee's effort enriches the manager to a smaller extent. As compared to all of these studies, our key innovation is to study pay-for-performance based on the principal's subjective evaluation, rather than on objective performance measures, in a dynamic model. Contrary to received wisdom based on static models, our dynamic model predicts that higher altruism sometimes lead to higher bonuses rather than lower, because altruism functions as a credibility device.

2.3 The model

We develop a principal-agent model where both players may be altruistic or spiteful toward each other. To focus on the effect of altruism and spite and avoid unnecessary complications, we assume that players are risk-neutral.[10] The utility of the principal is given by[11]

$$\Pi = \pi + \gamma U, \tag{2.1}$$

where $\pi \equiv q(e) - w$ is the principal's profits, w is the agent's compensation, and $q(e)$ is the value of output which is a function of agent's effort e. The term γU denotes the altruistic part of the principal's utility where U is the agent's utility and

[9] See also Fehr and Schmidt (2000) and Fehr et al. (2007) who study whether a fair-minded principal offers to a (potentially) fair-minded agent an explicit contract which imposes a fine for low performance, an implicit contract promising a bonus for high performance, or a trust contract consisting of a base salary only.

[10] See Casadesus-Masanell (2004) for a study of the interaction between agent's altruism and risk aversion.

[11] Unless otherwise stated, all variables are defined on a per-period basis.

γ denotes the degree of the principal's altruism. We call γ the principal's altruism parameter. Similarly, the agent's utility is given by

$$U = u + \alpha\Pi, \tag{2.2}$$

where $u \equiv w - \theta e$ is the agent's private utility, consisting of his compensation minus effort costs. The term $\alpha\Pi$ is the agent's altruistic utility, where α is the agent's altruism parameter. Clearly, an altruism parameter below zero reflects spite.

We assume that players know each other's altruism parameter. Further, we assume that neither player cares for the other's utility as much, or more, as he cares for his own utility; that is, $\alpha, \gamma \in (-1, 1)$. The players' utility functions – which are infinite recursions of each other – can therefore be expressed as[12]

$$\Pi = \frac{1}{1-\alpha\gamma}\left(\pi + \gamma u\right), \quad \text{and} \quad U = \frac{1}{1-\alpha\gamma}\left(u + \alpha\pi\right). \tag{2.3}$$

In each period that the agent is employed, he can exert low effort or high effort, denoted by $e \in \{0, 1\}$. Low effort results in low value of output, $q = L$, while high effort yields high value of output, $q = H$. The costs of exerting high effort to the agent are θ while exerting low effort is, for reasons of convenience but without loss of generality, assumed to be not costly. To make the problem interesting, we assume $H - \theta > L > 0$ and $H - \theta > \overline{\Pi} + \overline{U}$, where $\overline{\Pi} > 0$ and $\overline{U} > 0$ are the principal's and the agent's outside options, respectively. $\overline{\Pi}$ and $\overline{U}$ are independent of the players' type, which has two implications. First, the pair (α, γ) is specific to this relationship and thus does not reflect players' general altruistic or spiteful attitudes. Second, feelings of altruism or spite vanish if the players do not contract with each other (anymore). Lastly, note that $\overline{\Pi}$ and $\overline{U}$ can be interpreted as including the expected altruism or spite that would result when being matched to some other player.

The second important ingredient of the model is that both effort and output can only be assessed subjectively, and hence contracts that condition on effort or output are not enforceable in court. Any contract that includes compensation or

[12] For a thorough exposition of when interdependent utility functions form a unique system and can be defined on underlying utility drivers instead of some person's utility, see Bergstrom (1999).

punishment that is conditional on effort or output must therefore be self-enforcing. This can only hold in a repeated interaction setting. The model therefore includes an infinite number of periods, where players discount utility from future periods by a constant common factor δ. The self-enforceability of relational contracts depends further on the punishments inflicted upon the principal when she does not keep a promise or carry out a threat when the contract prescribes her to do so. Many models assume that the agent plays a trigger strategy that prescribes to break off trade when the principal reneges (e.g. Bull 1987, MacLeod and Malcomson 1989 and 1998, and Levin 2003). Following Baker et al. (1994), we do not impose that the agent always leaves the firm after contract breach. Instead we assume that after contract breach the agent will never believe any promise or threat in the future anymore.[13] Note that because this is a dynamic game of complete information, we solve for the subgame perfect equilibrium of the game.

The order of the game within a period is as follows. First, the principal chooses whether or not to hire the agent. Second, if the agent is hired, the players bargain over a contract which may contain a contractible base salary (s), a promise to pay a bonus (b) in case of high output, and a threat of dismissal in case of low output.[14] The outcome of contract negotiations is assumed to be given by the generalized Nash bargaining solution. We assume bargaining in order to create an environment in which it can be optimal to use a threat of dismissal as an incentive device. This is never the case if the principal holds all the bargaining power, since a threat of dismissal is only effective if the agent is paid costly rents (see MacLeod and Malcomson 1989). Assuming Nash bargaining thus enables us to study the impact of altruism and spite on the use of a threat of dismissal.[15] Third, if a contract

[13] One could argue that, if the principal employs more than one agent, she could overcome a possible credibility problem by using a tournament (see e.g. Carmichael 1983). While this may resolve the principal's problem of credibly committing to pay a bonus, multiple agents introduce a new credibility problem in the assignment of the bonus. That is, the principal will be tempted to award prizes to the agents she likes most rather than to the best-performing agents.

[14] Throughout the chapter, we abstract from the use of severance pay (or bonds). Severance pay leaves the credibility of bonus pay unaffected (MacLeod and Malcomson 1989). However, negative severance pay (that is, contracting on a transfer from the agent to the principal in case the principal dismisses the agent) can be used to make a threat of dismissal credible.

[15] The generalized Nash bargaining solution rests on three axioms: invariance, efficiency, and independence of irrelevant alternatives. The generalized Nash bargaining solution is a cooperative solution concept in the sense that players cooperate in dividing a given surplus from the labor

is agreed upon, the agent chooses an effort level. Fourth, output is realized and observed by both players. The principal decides whether or not to pay the bonus, and payoffs are realized. After that, a new period starts. By not rehiring the agent in this new period, the principal can exercise a threat of dismissal.

Typically, repeated games have many non-stationary equilibria. As is usual in the literature, we focus on stationary equilibria, that is, the principal and agent agree to the same contract every period.[16] Also, we assume that the contract only contains a particular type of incentive if necessary and credible.

2.4 Analysis

We start our analysis by deriving the conditions for the credibility of incentives. Next, we study under what conditions the agent exerts high effort. The section concludes with characterizing the optimal relational contract for any pair (α, γ).

2.4.1 Credibility of incentives

Before we can derive the credibility of relational incentives, we need to derive the value of the relational contract to the principal. The value of the relational contract to the principal depends, among others, on the outcome of the contract negotiations between her and the agent. Recall that we assume that the outcome of the negotiations is given by generalized Nash bargaining, which states that the agent's total

relationship between each other. However, generalized Nash bargaining does not overcome the central commitment issues studied in this chapter, namely the credibility of bonus pay and a threat of dismissal. Credibility of incentives may only exist by virtue of the agent's trigger strategy and repeated play. For an extensive discussion on how bargaining affects relational contracts, see Cordero Salas (2011).

[16] By Levin (2003)'s argument, stationary contracts can provide the agent with the same incentives as non-stationary compensation plans. This holds because in this model future rents and a bonus are perfect substitutes. Also note that the credibility of the bonus is unaffected by the choice whether to use future rents or a bonus (see MacLeod and Malcomson 1989). Finally, the credibility of a threat of dismissal is independent of the issue of (non-)stationarity, as will become clear in the next section. Hence, any complex non-stationary compensation plan can be replicated by a simple stationary one, maintaining the agent's incentives and the credibility of the contract.

compensation w is the solution to:

$$\max_{w} \Phi \equiv \left[\Pi(w) - \overline{\Pi}\right]^{1-\beta_A} \left[U(w) - \overline{U}\right]^{\beta_A}, \tag{2.4}$$

subject to the players' participation constraints

$$\Pi(w) - \overline{\Pi} \geq 0 \text{ and } U(w) - \overline{U} \geq 0. \tag{2.5}$$

If the players fail to reach an agreement, then they receive their outside option utilities. The agent's bargaining power is represented by $0 < \beta_A < 1$. Note that in the extreme case where $\beta_A = 0$, the principal has the power to make the agent a take-it-or-leave-it offer.

Solving (2.4) yields the agent's total compensation level given his effort choice:

$$w^*(e) = \frac{1}{1-\alpha}\left[(1-\alpha\gamma)\overline{U} + \theta e - \alpha q(e)\right] + \beta_A \frac{1-\alpha\gamma}{(1-\alpha)(1-\gamma)} S(e), \tag{2.6}$$

where $S(e) \equiv q(e) - \theta e - (1-\gamma)\overline{U} - (1-\alpha)\overline{\Pi}$ measures the surplus of the relationship. It is easy to verify that for any $\beta_A > 0$ and $S(e) > 0$ the agent earns a rent from the job. Also, the agent's compensation and thereby his utility, are increasing in the agent's bargaining power β_A, as in Cordero Salas (2011). Next, it can be shown that altruism gives rise to compensating wage differentials, as in Grund and Sliwka (2005), Delfgaauw and Dur (2007), and Arce (2010). However, the sign of these depend on the agent's bargaining power.[17]

[17] Differentiating $w^*(e)$ to α and γ yields:

$$\begin{aligned}
\frac{\partial w^*(e)}{\partial \alpha} &= -(1-\beta_A)\frac{1}{(1-\alpha)^2}\left[q(e) - \theta e - (1-\gamma)\overline{U}\right] + \beta_A \frac{\gamma}{1-\gamma}\overline{\Pi} \gtreqless 0 \text{ and} \\
\frac{\partial w^*(e)}{\partial \gamma} &= -(1-\beta_A)\frac{\alpha}{1-\alpha}\overline{U} + \beta_A \frac{1}{(1-\gamma)^2}\left[q(e) - \theta e - (1-\alpha)\overline{\Pi}\right] \gtreqless 0.
\end{aligned}$$

Note that in the extreme case where the agent has no bargaining power at all, it holds that $\frac{\partial w^*(e)}{\partial \alpha} < 0$. The intuition is that the principal extracts all of the agent's altruistic utility by decreasing his compensation. In the other extreme where $\beta_A = 1$, it is the agent who extracts all rents from the relationship. Whether the agent can extract more utility from the principal if he becomes more altruistic then depends on the principal's feelings. If the principal is spiteful, she dislikes working with a more altruistic agent (simply because the agent enjoys more utility when more altruistic) and so she must be compensated; hence $w^*(e)$ decreases in α if $\beta_A = 1$ and $\gamma < 0$. On the other hand, if the principal is altruistic, she enjoys employing a more altruistic agent and

Given that the agent's compensation equals $w^*(e)$, both players' participation constraints reduce to

$$S(e) \equiv q(e) - \theta e - (1-\gamma)\overline{U} - (1-\alpha)\overline{\Pi} \geq 0. \qquad (\text{PC}_P, \text{PC}_A)$$

The reason for why the players' participation constraints are identical is that players divide the surplus of the contract proportional to their bargaining power. Hence, all that matters for participation is whether the relationship yields a non-negative surplus. Also note that because both players have some bargaining power, each player's utility is increasing in $S(e)$. We are now ready to derive the credibility of relational incentives.

A threat of dismissal is only credible when the principal is willing to fire the agent if he is caught shirking. Recall that if a contract contains a threat of dismissal and the agent is nevertheless retained after shirking, he will never believe a threat or promise in the future anymore. Hence, if the agent is retained after he is caught shirking, the agent will forever shirk. Therefore, a threat of dismissal is credible only if the principal is better off dismissing the agent rather than retaining the agent knowing that he will shirk, $\overline{\Pi} \geq \Pi(w^*, e=0)$. This condition is the reverse of the participation constraint with $e = 0$:

$$S(e=0) \equiv L - (1-\gamma)\overline{U} - (1-\alpha)\overline{\Pi} \leq 0. \qquad (\text{CC}_D)$$

The above condition reveals that the surplus from the low-effort contract is increasing both in the principal's and the agent's altruism. The intuition is that altruism entails a social benefit from working together. Hence, the principal is not willing to fire the agent if altruism is sufficiently high. This is our first key result.

Proposition 2.1 *If* $\alpha \geq 1 - \frac{1}{\overline{\Pi}}\left[L - (1-\gamma)\overline{U}\right]$*, then a threat of dismissal is not credible.*

The other non-contractible part of the contract is a promise to pay a bonus for high output. For the bonus to be credible, it must be in the principal's interest to

so $w^*(e)$ increases in α if $\beta_A = 1$ and $\gamma > 0$. A similar but reverse intuition can be provided for the impact of the principal's altruism on the agent's compensation.

pay it conditional on observing high output. Adhering implies that the principal must incur bonus costs in the present, but she retains the opportunity to implement credible incentives in the future. Assuming a threat of dismissal is credible, one can derive that a bonus is credible if:

$$\frac{\delta}{1-\delta}\left[\Pi\left(s,b,e=1\right)-\overline{\Pi}\right]-\frac{1-\gamma}{1-\alpha\gamma}b\geq 0. \qquad (\mathrm{CC}_B)$$

Similarly, if the principal prefers to retain the agent after contract breach ($S\left(e=0\right)>0$), credibility of the bonus is guaranteed if

$$\frac{\delta}{1-\delta}\left[\Pi\left(s,b,e=1\right)-\Pi\left(w^{*},e=0\right)\right]-\frac{1-\gamma}{1-\alpha\gamma}b\geq 0. \qquad (\mathrm{CC}'_B)$$

The credibility constraints essentially state that paying a bonus is only credible when doing so yields sufficient rents in the future. Therefore, the credibility of a bonus increases in the principal's discount factor δ. Further, it can be easily derived that the credibility of the bonus increases in the agent's altruism. Assuming a threat of dismissal is credible, and the agent earns total compensation equal to $w^{*}\left(e=1\right)$ as is shown to be always the case in the next subsection, one can derive:

$$\frac{\partial \mathrm{CC}_B}{\partial\alpha}=\frac{\delta\left(1-\beta_A\right)}{1-\delta}\frac{1}{\left(1-\alpha\right)^2}\left[H-\theta-\left(1-\gamma\right)\overline{U}+\frac{\left(1-\alpha\right)^2\gamma}{1-\gamma}\overline{\Pi}\right]>0, \qquad (2.7)$$

where CC_B denotes the left-hand side of (CC_B), and the sign follows if participation is ensured. The intuition is the same as the one behind Proposition 2.1. If the agent is more altruistic the surplus of the relationship increases. Some of the increase of the surplus accrues to the principal depending on his bargaining power. It follows that it is more valuable for the principal to uphold the relational contract with the agent. The credibility of the bonus also increases in the principal's altruism:

$$\frac{\partial \mathrm{CC}_B}{\partial\gamma}=\frac{\delta\left(1-\beta_A\right)}{1-\delta}\frac{1}{\left(1-\gamma\right)^2}\left[H-\theta+\frac{\left(1-\gamma\right)^2\alpha}{1-\alpha}\overline{U}-\left(1-\alpha\right)\overline{\Pi}\right]>0. \qquad (2.8)$$

The reason is that, as above, an increase in the principal's altruism increases the

principal's rents from the relationship. In addition, a more altruistic principal experiences lower costs of paying a bonus, as she internalizes part of the benefit of the bonus to the agent. The same results hold if a threat of dismissal is not credible.[18] Hence, our second key result is:

Proposition 2.2 *Let $\overline{b}$ denote the maximum level of the bonus that the principal can credibly promise. It holds that $\overline{b}$ is strictly increasing in α and γ ($\frac{\partial \overline{b}}{\partial \alpha} > 0$ and $\frac{\partial \overline{b}}{\partial \gamma} > 0$).*

2.4.2 The optimal relational contract

Next consider the agent's incentive compatibility constraint. The agent is only willing to exert high effort if doing so yields higher expected lifetime utility than shirking does. Let's first suppose the contract contains a credible threat of dismissal. The incentive compatibility constraint is then given by:

$$\frac{1-\alpha}{1-\alpha\gamma}b + \frac{\delta}{1-\delta}\left[U\left(s,b,e=1\right) - \overline{U}\right] + \frac{\alpha}{1-\alpha\gamma}\left(H-L\right) \geq \frac{1}{1-\alpha\gamma}\theta. \tag{2.9}$$

The right-hand side denotes the agent's costs of exerting high effort. The left-hand side describes the benefits from exerting high effort which may be threefold: a bonus, a stream of future rents, and intrinsic joy of enriching the principal arising from the agent's altruism. Assuming that the agent's compensation equals $s + b = w^*\left(e=1\right)$ as described by (2.6) (which we prove always to be the case in the proof of Proposition

[18] Differentiating the left-hand side of $\mathrm{CC}_B^{'}$ yields:

$$\begin{aligned}\frac{\partial \mathrm{CC}_B^{'}}{\partial \alpha} &= \frac{\delta\left(1-\beta_A\right)}{1-\delta}\frac{1}{\left(1-\alpha\right)^2}\left(H-\theta-L\right) > 0, \text{ and}\\ \frac{\partial \mathrm{CC}_B^{'}}{\partial \gamma} &= \frac{\delta\left(1-\beta_A\right)}{1-\delta}\frac{1}{\left(1-\gamma\right)^2}\left(H-\theta-L\right) > 0.\end{aligned}$$

The intuition behind these comparative static results is slightly different from the one in the main text. The reason is that the principal's utility from reneging is not her outside option utility, but the utility from the low-effort contract. It always holds that the principal's utility from a contract increases in both players' altruism. However, the principal's utility rises faster with altruism under the high-effort contract compared to the low-effort contract. Therefore, higher altruism makes it more attractive to the principal to adhere to the bonus also if a threat of dismissal is not credible.

2.4), the minimum required bonus to induce high effort can be written as:

$$b^* = \frac{1}{1-\alpha}\left[\theta - \alpha\left(H-L\right) - \frac{\delta\beta_A}{1-\delta}\frac{1-\alpha\gamma}{1-\gamma}S\left(e=1\right)\right], \tag{IC}$$

where $S\left(e=1\right)$ is described by (PC_P, PC_A) with $q(e)=H$ and $e=1$. It is easy to verify that the minimum required bonus b^* is decreasing in both players' altruism. The reason is twofold. First, a more altruistic agent experiences more altruistic utility from enriching the principal. Second, as derived in the previous subsection, the contract becomes more valuable to the agent if either player becomes more altruistic. Hence, provided that the contract contains a credible threat of dismissal, a lower bonus is required.

Next suppose that a threat of dismissal is not credible and, hence, is not part of the contract. Then, the agent's effort choice does not affect his future employment status. The reason is that the principal will retain the agent in any case. Consequently, the agent only takes the current costs and benefits of effort into account and exerts high effort if the bonus equals at least:

$$b^{**} = \frac{1}{1-\alpha}\left[\theta - \alpha\left(H-L\right)\right]. \tag{IC'}$$

This expression reveals that for sufficiently high agent's altruism, incentives are unnecessary to induce high effort. This is the case when $\alpha \geq \frac{\theta}{H-L}$. It follows directly that the optimal relational contract in this case is simply a flat wage contract without threat of dismissal. By generalized Nash bargaining, the agent's total compensation equals $w^*\left(e=1\right)$ as described by (2.6). Our next Proposition follows.

Proposition 2.3 *Sufficiently altruistic agents ($\alpha \geq \frac{\theta}{H-L}$) exert high effort even in the absence of incentives. In this case, the optimal relational contract is a contract without incentives that results in high effort. The agent's total compensation equals $s = w^*\left(e=1\right)$ as described by (2.6).*

Next we derive the optimal relational contract assuming that the agent needs incentives to exert high effort. The results are given in Proposition 2.4 and derived and explained in detail in the proof.

Proposition 2.4 *If the agent needs incentives to exert high effort ($\alpha < \frac{\theta}{H-L}$), the optimal relational contract is described by the following Table:*

	ToD credible	*ToD not credible*
Bonus credible	*a) ToD, $b \geq b^*$, $e = 1$* *b) ToD, flat wage, $e = 1$* *c) no ToD, $b \geq b^{**}$, $e = 1$*	*No ToD, $b \geq b^{**}$, $e = 1$*
Bonus not credible	*No contract*	*No incentives, $e = 0$*

where ToD denotes threat of dismissal. Cases (b) and (c) may apply in addition to case (a) only if respectively the agent's and the principal's future rents are sufficiently large (see condition (2.10) and (2.11) below). Further, the agent's total compensation always equals $s + b = w^*(e)$ as described by (2.6).

Proof. First, in the case where neither credibility constraint binds, both types of incentives are credible and so the contract can always induce high effort. By generalized Nash bargaining the agent's compensation is given by $s + b = w^*(e = 1)$ as described by (2.6) which yields the agent a rent. Given that the agent earns a rent, any bonus $b \geq b^*$ as described by (IC) combined with a threat of dismissal ensures high effort (Case (a)). However, there are potentially two alternatives, depending on the size of the players' rents. If the value of the agent's future rents is sufficiently high, then a threat of dismissal alone can induce high effort (Case (b)). This holds when:

$$\frac{\delta}{1-\delta}\left[U(w^*, e=1) - \overline{U}\right] \geq \frac{1}{1-\alpha\gamma}\left[\theta - \alpha(H-L)\right]. \tag{2.10}$$

If the value of the principal's future rents is sufficiently high, a bonus contract with $b \geq b^{**}$ and no threat of dismissal can also induce high effort (Case (c)). This holds when:

$$\frac{\delta}{1-\delta}\left[\Pi(w^*, e=1) - \overline{\Pi}\right] \geq \frac{1-\gamma}{(1-\alpha\gamma)(1-\alpha)}\left[\theta - \alpha(H-L)\right]. \tag{2.11}$$

Hence, if either condition (2.10) or (2.11) holds, the players have various contract

types at their disposal to induce high effort. It is easy to verify that both conditions hold for the same values of α and γ if $\beta_A = \frac{1}{2}$. However, if $\beta_A < \frac{1}{2}$ condition (2.11) binds for lower values of α and γ, whereas otherwise condition (2.10) binds for lower values of α and γ.

Second, consider the case where promising a bonus is credible while threatening with dismissal is not. Then, it is still possible to contract on high effort because the required bonus can be credibly promised. By Nash bargaining, the agent's compensation is given by $s + b = w^*(e = 1)$, where it must hold that $b \geq b^{**}$ as described by (IC').

Third, if neither incentive is credible, then it is not possible to induce high effort. However, if the threat of dismissal is not credible, it is by definition attractive to hire the agent even if he shirks. The agent is paid compensation resulting from generalized Nash bargaining, $s = w^*(e = 0)$ as described by (2.6). Hence, a flat wage contract without threat of dismissal is agreed to when neither incentive is credible.

The final case is where a threat of dismissal is credible but the minimum required bonus, b^*, is not. Since the threat of dismissal is credible, it is by definition not attractive to contract on low effort. An argument similar to that of MacLeod and Malcomson (1989) proves that the principal does not find it attractive to contract on high effort either, and so no contract is agreed to at all. The argument runs as follows.

The credibility constraint for bonus pay, given that the threat of dismissal is credible, can be rewritten as

$$b \leq \overline{b} \equiv \delta \left(\frac{1}{1-\gamma} (H - \gamma\theta) - \frac{1-\alpha\gamma}{1-\gamma} \overline{\Pi} - s \right), \qquad (\mathrm{CC}_B)$$

where $\overline{b}$ is the maximum credible bonus. Since $\overline{b} < b^*$, the maximum credible bonus $\overline{b}$ is too low to satisfy the incentive compatibility constraint (IC). The only way to satisfy (IC) then is to increase the agent's rents by raising the agent's total compensation above $w^*(e = 1)$. Even though this will result in a division of rents that is different from the division implied by the generalized Nash bargaining solution,

it may still make the principal better off compared to the case where no contract is agreed to. However, the following argument shows that this is not the case. First note that the impact of increasing the base salary on the incentive compatibility constraint (IC) is zero if (CC_B) is met with equality:

$$\frac{\partial \text{IC}}{\partial s} = \frac{1-\alpha}{1-\alpha\gamma}\frac{\partial \bar{b}}{\partial s} + \frac{\delta}{1-\delta}\frac{1-\alpha}{1-\alpha\gamma}\left(1 + \frac{\partial \bar{b}}{\partial s}\right) = 0, \tag{2.12}$$

where, following (CC_B), $\partial \bar{b}/\partial s = -\delta$. The reason is that increasing the base salary makes the contract less profitable to the principal, and hence the maximum credible bonus decreases. Therefore net incentives from raising the base salary do not change until the maximum credible bonus equals zero. Further, if the principal can only credibly promise a bonus that is equal to zero, it must hold that she earns no rent from the contract, $\Pi\left(s, b = 0, e = 1\right) = \overline{\Pi}$. Moreover, the principal must still increase the base salary in order to satisfy the incentive compatibility constraint (IC). But this violates the principal's participation constraint and so she will not enter into a high-effort contract. ■

2.4.3 Altruism and contract selection

So far we have found that players may agree to different types of relational contracts. In this subsection we investigate what type of contract is selected for any pair (α, γ). Naturally there are many different contract mappings, depending on the exact parameterization of the model. However, there is one important qualitative distinction between the different mappings: either higher altruism always increases the surplus from the relationship, or it may reduce it sometimes. We start with illustrating the first case in Figure 2.1, which has the following properties.

First, altruism must be sufficiently high to satisfy the players' participation constraints. Naturally, PC($e = 0$) lies to the right of PC($e = 1$) as output is lower in the low-effort contract.

Second, high effort is only sustainable if neither player wants to deviate from a high-effort contract. Above the line $\alpha = \frac{\theta}{H-L}$ the agent has sufficiently high altruism to exert high effort without incentives. Hence, above this line the optimal

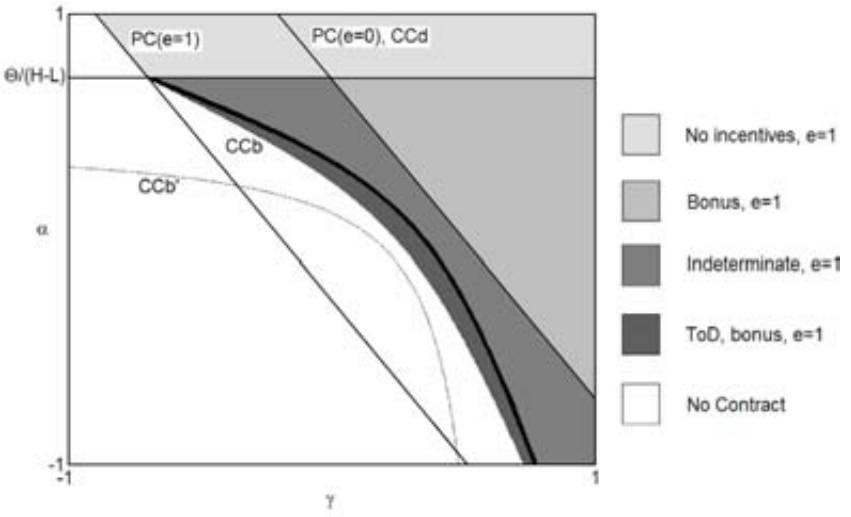

Figure 2.1: Contract selection as a function of α and γ. $H = 2.1$, $L = 1$, $\theta = 0.7$, $\overline{U} = 0.58$, $\overline{\Pi} = 0.41$, $\beta_A = 0.3$, and $\delta = 0.75$.

contract is a contract without incentives that results in high output (Proposition 2.3). Below the line $\alpha = \frac{\theta}{H-L}$, incentives are required to induce high effort, which must be credible. PC($e = 0$) demarcates whether the principal finds it attractive to hire the agent if he shirks. Therefore to the right of PC($e = 0$) altruism is too high to make credible use of a threat of dismissal (Proposition 2.1). Next, depending on the credibility of a threat of dismissal, the principal must promise the agent a bonus equal to at least b^* or b^{**} to induce high effort. Hence, if a threat of dismissal is credible, the credibility constraint for bonus pay becomes

$$\frac{\delta}{1-\delta}\left[\Pi\left(w^*, e=1\right) - \overline{\Pi}\right] - \frac{1-\gamma}{1-\alpha\gamma} b^* \geq 0, \qquad (\mathrm{CC}_B)$$

whereas otherwise it is

$$\frac{\delta}{1-\delta}\left[\Pi\left(w^*, e=1\right) - \Pi\left(w^*, e=0\right)\right] - \frac{1-\gamma}{1-\alpha\gamma} b^{**} \geq 0. \qquad (\mathrm{CC}'_B)$$

The reason that CC_B and CC'_B are evaluated at b^* and b^{**} is that players can always lower the bonus to this level without changing total compensation, which

would be optimal if a higher bonus is not credible. Hence the constraints denote when promising some bonus that induces high effort is feasible. Also note that whenever the agent needs incentives to exert high effort, CC_B lies strictly to the right of PC($e = 1$). The reason is that the principal requires a rent to be able to credibly promise a bonus.

Third, at CC_B the relationship is exactly that valuable such that the principal can promise the bonus b^* which, in combination with a threat of dismissal, induces high effort from the agent (Proposition 2.4, case a). As we move to the north-east, altruism becomes higher, and so both players' rents increase. Therefore, at some point it holds that either the principal's or the agent's rents are equal to the agent's costs of effort. This point is determined by either equation (2.10) or (2.11) and graphically represented by the fat line. If the agent's rents equal his cost of effort, an efficiency wage contract without bonus pay induces high effort (Proposition 2.4, case b). Alternatively, if the principal earns sufficiently high rents, she can credibly promise a bonus that alone induces high effort (Proposition 2.4, case c). As it is also possible to use both tools instead of only one, several contract types are optimal beyond the fat line till CC_D. The model's predictions concerning contract type are therefore indeterminate in this region.[19]

Fourth, anywhere beyond CC_D, a threat of dismissal is not credible, but the bonus b^{**} is credible and so the optimal contract is a bonus contract that induces high effort. However, for different parameterizations, b^{**} may not always be credible beyond CC_D as we will see in Figure 2.2.

Figure 2.2 again represents the mapping from (α, γ) pairs to contract type, but now for a lower value of the discount rate. This shifts the bonus credibility constraints to the right as higher altruism is required to make the bonus credible. The crucial difference between Figures 2.1 and 2.2 is that there is now an intermediate region where neither incentive is credible.[20] By Proposition 2.4, players then agree to a contract without incentives that results in low effort.

To understand why the principal loses the possibility to promise a bonus for some

[19] In Figure 2.1 (and Figure 2.2) the fat line represents equation (2.11). So, directly to the right of the fat line players can choose to only use a bonus.

[20] Such a region also emerges for a sufficiently high value of the agent's bargaining power (β_A).

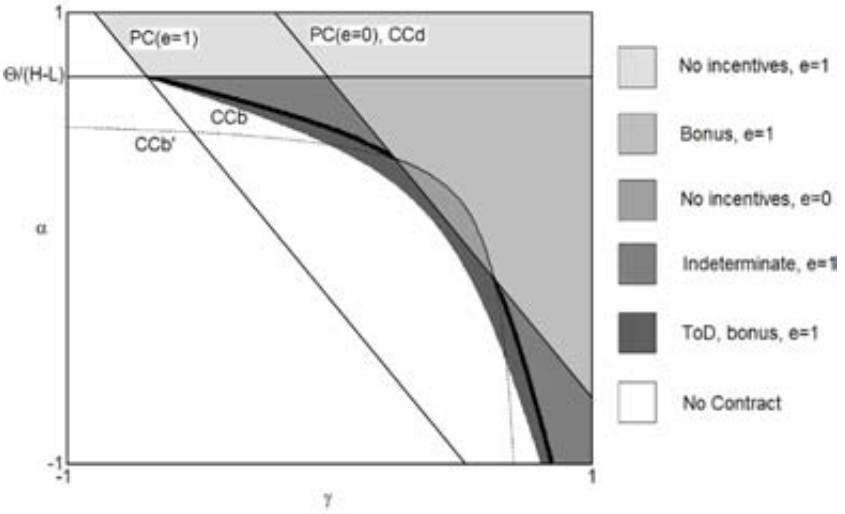

Figure 2.2: Contract selection as a function of α and γ. Same parameterization as in Figure 2.1, except $\delta = 0.6$.

pairs (α, γ) beyond CC_D, consider the following. If a threat of dismissal is no longer credible, this reinforces the credibility problem concerning the bonus. The reason is twofold. First, the principal's fallback utility from reneging is no longer her outside option utility but her utility from a low-effort contract. By definition the latter is weakly higher than the former if a threat of dismissal is not credible, and so the principal has weaker incentives to stick to her promise to pay a bonus. Second, if a threat of dismissal is not credible, the agent's rents cannot be used as an incentive device and so the minimum bonus must increase from b^* to b^{**}. This again gives the principal stronger incentives to withhold the bonus. For high values of δ it is never a problem that the bonus must increase from b^* to b^{**} because the present value of adhering to the principal is still large enough, as in Figure 2.1. However, for low values of δ the principal's present value of adhering is low, and so a discrete increase in the bonus has a relatively large impact on the credibility constraint. Hence the non-credibility of a threat of dismissal may put too much strain on the promise to pay a bonus.

The optimality of a low-effort contract for intermediate levels of altruism implies

two counterintuitive comparative static results. First, at the credibility constraint of a threat of dismissal, a marginal improvement in altruism is harmful to productivity and players' utility in equilibrium. The reason is that, at CC_D, a marginal increase in altruism renders a threat of dismissal non-credible, and hence, players lose the opportunity to credibly contract upon high effort. Second, at the credibility constraint for bonus pay (CC'_B), a marginal increase in altruism does not lead to lower bonuses as commonly found in the literature (Casadesus-Masanell 2004, Sliwka 2007, Shchetinin 2010, and Non 2012), but rather higher. The reason is that higher altruism enables credible use of a bonus. We summarize these results in our final Proposition.

Proposition 2.5 *Depending on the parameterization of the model, a region may exist where players agree on a low-effort contract without incentives. Starting from this low-effort contract, it may hold that for lower values of α and/or γ, effort and players' utility are strictly higher, while higher values of α and/or γ can lead to strictly higher bonuses.*

2.5 Practical and testable implications

The theoretical analysis conducted above yields a number of clear implications for management. First, we have shown that the credibility of relational incentives depends on the altruism or spite between a manager and employee. Altruism reinforces the credibility of bonus pay whereas it undermines the credibility of a threat of dismissal. Our model therefore suggests that firms in which relationships between management and employees are characterized by altruism benefit more from using bonus pay compared to efficiency wages. Further, when actively shaping organizational culture, firms should be aware that altruism does not necessarily improve the credibility of incentive schemes. This is only the case if incentive schemes rely sufficiently on promises rather than threats.

The second insight is that moderate altruism may lead to inefficiently low effort levels by employees. This is the case when altruism is too high to credibly threaten employees with dismissal, but too low to credibly promise a bonus. In response to

such a situation, firms can raise productivity by either stimulating or suppressing altruism. The reason is that doing the former makes bonus pay more credible, whereas the latter renders a threat of dismissal more credible.

Third, the presence of altruism increases employees' willingness to work for the firm in the first place. Therefore, altruism between management and employees not only affects employee motivation as described above, but also determines the total compensation level employees require to be willing to work for the firm (see also Grund and Sliwka 2005, Delfgaauw and Dur 2007, and Arce 2010).

The model's predictions can be tested using laboratory experiments and field data. In laboratory experiments, the researcher has the freedom to fine-tune the profitability of the various types of contracts, which is hard outside the lab. Moreover, in contrast to the field, one can easily observe when players decide not to enter into a contract. A bottleneck may be to bring feelings of altruism and, particularly, feelings of spite into the lab. This difficulty may be resolved by recruiting subject pools in which these feelings are already present. One could think of members of (non-)rivalrous sports clubs or student associations. Using field data, one can test whether there is a relationship between contract types and the degree of altruism between manager and employee.

There are avenues for future theoretical research as well. First, it would be interesting to endogenize feelings of altruism. Dur (2009), Englmaier and Leider (2012), and Non (2012) have taken some steps in this direction in one-shot games. Second, it is worth studying how our results change when performance evaluations are done by a potentially altruistic middle-manager who is not residual claimant. In one-shot games, this case has been studied by Prendergast and Topel (1996), Lee and Persson (2011), Giebe and Gürtler (2012), and Grund and Przemeck (2012). The next chapter of this thesis studies this case in a dynamic setting.

Chapter 3

Leniency Bias in Long-Term Workplace Relationships

3.1 Introduction

Many firms incentivize their employees on the basis of subjective performance evaluation (MacLeod and Malcomson 1998, Prendergast 1999, MacLeod and Parent 2000, Gibbs et al. 2004, Gibbs 2012). Moreover, the actual appraisal of employee performance is often not conducted by firm owners, but delegated to supervisors. There is ample evidence that supervisors exert bias in appraising performance.[1] Such bias may be harmful to firms because it weakens the link between employees' actual performance and their performance evaluation. In turn, this reduces employees' incentives to exert effort. Firms therefore have an interest in stimulating supervisors to report performance truthfully. This chapter develops and analyzes a model to study how firms can achieve this through designing optimal contracts for supervisors and employees.

In the model, the agent chooses a privately observable, continuous effort level that generates a high or low performance. To induce the agent to exert effort, the principal wishes to pay the agent a bonus if his performance is high. The

[1] Well-known biases are the leniency bias and the centrality bias. See for example Medoff and Abraham (1980), Jawahar and Williams (1997), Prendergast (1999), Moers (2005), Berger et al. (2011), and Bol (2011).

agent's performance is not verifiable, but can be subjectively evaluated. Since the principal does not observe the agent's performance herself, she hires a supervisor to perform this task. The supervisor is either selfish or altruistic towards the agent, and may therefore bias performance evaluations. Supervisor's altruism is observed by the agent but not by the principal. For simplicity, I focus on the case where the unconditional probability that the supervisor is altruistic equals one half.

The agency problem is studied in an infinitely repeated game. As is usual in the repeated games literature, the agent is assumed to play a trigger strategy that prescribes to shirk in all future periods if the supervisor has once biased his performance evaluation (see e.g. Bull 1987). Next, as in Baker et al. (1994), I assume there is a verifiable signal about the agent's performance available to the principal. The verifiable signal is necessary for incentivizing the supervisor to report performance truthfully, as will be explained below. Allowing for objective incentives for the agent on the basis of this verifiable signal does not change the results of the analysis qualitatively, and so for simplicity I abstract from this in the analysis. Further, I assume there is a large pool of supervisors and agents. Therefore, if a current labor relationship breaks down, the principal will be rematched to a new agent and/or supervisor in the next period. Finally, throughout the chapter I restrict attention to the case where the principal offers a uniform contract to the supervisor.

The analysis yields two key results. First, given the optimal contracts, supervisor's altruism may provide a net incentive to report performance truthfully, rather than to bias evaluations. Second, for relatively low and relatively high values of the discount factor, the principal optimally designs the supervisor's contract such that it only attracts altruistic types. In the remainder of the Introduction, the results are described and explained in more detail.

In case the supervisor is altruistic, he enjoys to some extent the well-being of the agent. Therefore, the altruistic supervisor has an incentive to report the agent's performance to be high, irrespective of the agent's actual performance. The reason is that reporting high performance will yield the agent a valuable bonus. However, the agent only has an incentive to exert effort if he expects the supervisor to report performance truthfully. In the repeated game, the supervisor will do so only

if the following holds: reporting performance truthfully yields future rents, the discounted value of which exceeds the current benefit from overreporting. To satisfy this condition in the most efficient way, the principal optimally uses the following three compensation elements.

First, note that the lower the altruistic supervisor's current benefit from overreporting, the lower the rents he requires to report performance truthfully. The principal therefore optimally makes it costly to the supervisor to report high performance. This is achieved by tying the supervisor's compensation to his evaluation report, as in Giebe and Gürtler (2012). Clearly, a monetary penalty following a high performance rating makes it attractive for the selfish supervisor type to bias performance ratings downward. This is costly to the principal since now the selfish type must be paid rents to report performance truthfully. The optimal penalty for reporting high performance therefore trades off the rents paid to the selfish type against the rents paid to the altruistic type, and depends on whether or not the principal screens for a supervisor type, as will become clear below.

Second, to ensure that reporting performance truthfully yields future rents, the supervisor receives a monetary reward if the verifiable signal about the agent's performance is high. To understand why, note that the agent responds to a biased performance evaluation by shirking in the future. If the agent shirks, the verifiable signal will be low. Therefore, biasing the agent's performance evaluation reduces the supervisor's compensation in the future.

Third, to maximize the supervisor's incentives to report performance truthfully, his utility after biasing a performance evaluation is set as low as possible. This is achieved by paying the supervisor a low base salary. The base salary can be set as low as necessary. This holds because a low base salary can be compensated for by paying the supervisor a higher reward if the verifiable signal about the agent's performance is high. Because the supervisor can always take his outside option, the implication is that his utility after biasing an evaluation optimally equals at most his reservation utility.

In equilibrium, the principal sets the three compensation elements discussed above such that the supervisor's uniform contract is attractive to both types, or

only to the altruistic type. In the former case, I say that the supervisor's contract is a *pooling contract*, whereas in the latter case I say that the supervisor's contract is a *screening contract*. The crucial feature of the screening contract is that it reveals the supervisor's type. This information is valuable to the principal because it enables her to eliminate the supervisor's current benefit from biasing the performance evaluation. This is achieved by setting the right penalty for giving the agent a high performance evaluation. The benefit to the principal is that the supervisor does not require a rent to report performance truthfully. Further, since in this case inducing correct performance evaluations from the supervisor does not generate agency costs, it is also optimal to provide the agent the first-best level of the bonus. Screening for one supervisor type thus yields the principal first-best profits, conditional on being matched. Assuming that the principal optimally screens, it holds that she optimally screens for the altruistic type. The reason is that, in contrast to the selfish type, the altruistic type enjoys working with the agent. The altruistic type therefore demands a lower total compensation level to accept the job. This feature also explains why the screening contract does not attract the selfish type.

Next, if the principal offers the pooling contract to the supervisor, at least one supervisor type enjoys a benefit from biasing the performance evaluation in the current period. This is inevitable because one type is selfish and the other is altruistic, but the penalty for giving a high performance evaluation can only take one value. It follows that at least one type must enjoy future rents to be willing to report performance truthfully. Interestingly, the altruistic supervisor enjoys future rents from reporting truthfully precisely because he is altruistic. The reason is that he enjoys altruistic utility from working with the agent. This utility cannot be extracted from the altruistic type, since doing so would make the job unattractive to the selfish type. It follows that, as long as he reports performance truthfully, the altruistic type's utility is strictly above his reservation utility. However, as explained above, the optimal contract ensures that any type of supervisor earns at most his reservation utility after biasing a performance evaluation. Biasing an evaluation thus implies that the altruistic type will lose the future rents derived from working with the agent. Therefore, supervisor altruism not only gives an incentive to overreport performance, but

it also provides an incentive to report performance truthfully. Moreover, if the supervisor is sufficiently patient, the latter effect outweighs the former. In that case, the net incentive from supervisor altruism is to report performance truthfully, rather than to bias performance evaluations.

The optimal contracts in case the principal attracts both supervisor types are as follows. If the supervisor is very patient, then even if the agent's bonus is high, the altruistic type prefers to report performance truthfully, as described above. The principal therefore only needs to incentivize the selfish type to report performance truthfully. By the same intuition as in the screening case, the optimal way to do so is by letting the supervisor's compensation be independent of the supervisor's evaluation report. This ensures that the selfish type is indifferent to his evaluation report, and hence does not have to be paid costly rents for rating truthfully. As in the screening case, agency costs from subjective performance evaluation are now fully avoided, implying it is optimal to provide the agent with the first-best level of the bonus.

If the supervisor is less patient, and the agent's bonus is high, the altruistic supervisor prefers to bias performance evaluations. The principal can respond to this situation by reducing the agent's bonus, such that supervisor altruism again provides a net incentive to report performance truthfully. Alternatively, she may increase the supervisor's compensation to ensure reporting truthfully yields sufficient rents. The principal can also pursue both strategies simultaneously. However, in any case it holds that the lower the supervisor's patience, the lower the principal's profits. Therefore, if the supervisor is sufficiently impatient, it becomes optimal to screen. Screening is costly because the supervisor may reject the job offer, in which case the principal earns his reservation utility. Yet, these costs are outweighed by the benefits from increasing the agent's incentives, and eventually reducing the supervisor's compensation, conditional on the supervisor accepting his contract. I call this the 'incentive motive' for screening.

Finally, screening is also optimal if players are sufficiently patient. This may seem counterintuitive, since it has just been explained that attracting both supervisor types yields high profits in this case. However, if the principal does not screen,

she cannot enjoy the compensating wage differential that hiring only the altruistic supervisor gives rise to. If the principal is sufficiently patient, she is willing to incur the costs of screening to enjoy higher profits once having found an altruistic supervisor. I call this the 'wage differential motive' for screening.

The chapter is organized as follows. Section 3.2 discusses the related literature. Section 3.3 describes the model, which is solved in section 3.4. Section 3.5 finishes with concluding remarks.

3.2 Related literature

Supervisor altruism as a source of bias in performance evaluation has received considerable attention. A seminal paper in the field is Prendergast and Topel (1996). The authors show that in response to biased performance ratings, firms may optimally reduce incentives for agents, limit authority of supervisors, and use bureaucratic rules in pay and promotion decisions. Lee and Persson (2011) extend Prendergast and Topel (1996) by allowing the agent and the supervisor to be mutually altruistic. They show that supervisor's altruism induces leniency, whereas agent's altruism induces loyalty. Moreover, the benefits from the agent's loyalty may outweigh the negative effects from the supervisor's leniency, such that the principal is better off with good social relationships in the workplace. Another extension of Prendergast and Topel (1996) is Grund and Przemeck (2012), who consider a situation where two inequality averse agents must be supervised. In this case, the altruistic supervisor not only inflates ratings, but also compresses ratings in order to reduce inequality between the agents.[2]

As already noted by Prendergast and Topel (1996), incentives can be used to align supervisor's behavior with the firm's interest, rather than with his social pref-

[2]Other supervisor preferences potentially causing distorted ratings have been analyzed in the economic literature. Müller and Weinschenk (2011) derive optimal contracts when the supervisor suffers from the 'horns and halo' effect, that is, the tendency to give performance evaluations close to the ones given in the past. Golman and Bhatia (2012) assume the supervisor observes noisy signals of the agent's performance, and feels worse about unfavorable mistakes compared to favorable ones. Kamphorst and Swank (2012) analyze supervisor's rating behavior when the supervisor simultaneously wants to strengthen the agent's self-confidence, and come across as a capable evaluator of the agent's performance.

erences. This claim receives empirical support from Bandiera et al. (2009). In a field experiment, the authors exogenously vary supervisors' compensation from a fixed wage to a bonus based on the productivity of supervised workers. Bandiera et al. (2009) report that supervisors respond to this change by assisting workers in a more firm-efficient way, rather than assisting the workers they are socially connected to. In line with this, Berger et al. (2011) find that supervisors who share in company profits or receive performance pay make better promotion decisions. Studying this issue theoretically, Giebe and Gürtler (2012) derive optimal incentive contracts for employees and supervisors when it is the supervisor's task to evaluate the agent's performance. In their model, the supervisor may be either selfish or altruistic towards the agent, which is private information as in Prendergast and Topel (1996). The authors show that inducing correct performance evaluations from the altruistic type may only be possible at the cost of reducing the agent's incentives, and paying the supervisor costly rents. For this reason, the principal may optimally choose not to eliminate leniency from the altruistic supervisor at all. In this case the agent will still exert effort because with some probability she will face the selfish supervisor, who reports performance truthfully.

The present chapter contributes to the existing literature in the following ways. First, the model developed and analyzed here is dynamic which enables me to study workplace relationships where interaction is repeated over time. This feature of the model gives rise to the first key result, namely that supervisor altruism may provide a net incentive to report performance truthfully, rather than to bias performance evaluations. Second, as is realistic in long-term workplace relationships, I assume that the agent knows the supervisor's type, and therefore knows whether or not the supervisor will be lenient in rating his performance. Last, this chapter highlights an optimal response to the possibility of supervisor bias in performance evaluations that has not been considered before, namely to attract only one supervisor type. A screening contract reveals the supervisor's type, which consequently enables the principal to more efficiently incentivize the supervisor to report performance truthfully. To the best of my knowledge, this is the first study to analyze optimal screening for

supervisor types.[3]

By studying an infinitely repeated game, this chapter is also related to the literature on relational contracts (see e.g. Bull 1987, MacLeod and Malcomson 1989 and 1998, Baker et al. 1994, and Levin 2003). Closely related are Cordero Salas and Roe (2012) and the previous chapter of this thesis. Both studies show that a more altruistic principal finds it easier to credibly promise a bonus to the agent. Similarly, the analysis here shows that supervisor altruism may provide an incentive to report performance truthfully. In contrast to the present chapter, the relational contracts literature generally assumes that a residual claimant principal evaluates the agent's performance. An exception is Thiele (2013), who studies a principal-supervisor-agent-model that allows for the possibility of collusion between the supervisor and the agent. This analysis differs from Thiele (2013) in that I abstract from collusion. Instead, it is assumed that bias in performance evaluations stems only from the supervisor's altruism towards the agent. Also, I study the role of uncertainty about the supervisor's type to the principal, and assume there is verifiable information about the agent's performance available to the principal.

Finally, Sol (2010) studies the use of peer evaluations to incentivize employees, and allows peers to be altruistic or spiteful towards each other. Altruism and spite lead peers to internalize part of their co-workers' well-being, which implies that peers have an incentive to bias performance evaluations. Sol (2010) shows that the principal can ensure truthful peer evaluations by reducing the bonus tied to a positive evaluation, which is sometimes found to be optimal in the present analysis as well. Sol (2010) also shows that if peer evaluation becomes severely constrained because feelings of altruism or spite are strong, the principal optimally uses team incentives in addition to peer evaluation.

[3] Since the seminal analysis by Spence (1973), screening for productive agents has become a well-known concept. Recently, screening for agents with some social preference also has attracted quite some attention. See for example Sliwka (2007) (conformism and altruism), Delfgaauw and Dur (2007) (intrinsic motivation), Delfgaauw and Dur (2008) (public service motivation), Kosfeld and Von Siemens (2010, 2011) and Non (2012) (reciprocity), and Von Siemens (2011, 2012) (inequality aversion).

3.3 The model

Consider a world with one principal, many agents and many supervisors. All players live an infinite number of periods, denoted by $t = 0, 1, 2, ...$ Future periods are discounted at a common discount factor δ. Each period the principal may employ one agent and one supervisor. When employed by the principal, the agent chooses a privately observable effort level, e. Effort yields a non-verifiable performance, denoted by $V \in \{0, 1\}$. The probability that the agent's performance is high is $\Pr[V = 1] = e$. V is observed by the agent and the supervisor, but not by the principal.[4] The agent's effort also produces a verifiable signal $z \in \{0, 1\}$, which is modeled in the same way as in Baker et al. (1994). The probability that the verifiable signal takes a high value is $\Pr[z = 1] = \mu e$. The parameter μ has an expectation $E[\mu] = 1$, a variance $\sigma_\mu^2 > 0$, and μ is privately observed by the agent before exerting effort. The costs of exerting effort are given by $\frac{1}{2}\theta e^2$. Note that the first-best effort level is therefore equal to $e^{FB} = \frac{1}{\theta}$. To ensure that the probability that $V = 1$ is properly defined, it is assumed that $\theta \geq 1$. It is also assumed that θ and the support of μ are such that $\Pr[z = 1] \in [0, 1]$.

In the analysis, I assume that the principal incentivizes the agent only on the basis of subjective evaluation of the agent's performance V.[5] Since the principal does not observe the agent's performance herself, it is the supervisor's task to evaluate the agent's performance. The supervisor's evaluation report is denoted by $d \in \{0, 1\}$, where $d = 0$ means "$V = 0$", and $d = 1$ means "$V = 1$". Importantly, the

[4] It is a feature of many large firms that firm-owners do not observe the performance of individual employees. See also e.g. Prendergast and Topel (1996) and Giebe and Gürtler (2012), who make the same assumption.

[5] Sometimes the principal can attain strictly higher profits by also incentivizing the agent on the basis of the verifiable signal z (namely in cases II and III of Proposition 3.2). However, if σ_μ^2 is sufficiently large but finite, the key results of the chapter are qualitatively the same as in the case where the principal only uses subjective incentives for the agent. A formal proof of this claim is provided in Appendix 3.B. The intuition behind it is as follows (also see Baker et al. 1994). If the agent is incentivized on the basis of the verifiable signal z, he will vary his effort level with the realization of the parameter μ. However, variation in the agent's effort level is costly because the agent's effort costs are convex. For these costs the agent must be compensated through his wage. Therefore, if σ_μ^2 is sufficiently high, using objective incentives is so costly such that it is always optimal to also include subjective incentives in the agent's contract, in which case all results of the chapter hold.

principal can commit to reward the agent on the basis of the supervisor's report.[6] The principal can therefore offer the agent a fully enforceable contract, which consists of a base salary a, and a bonus b to be paid if the supervisor reports performance is high ($d = 1$). The agent's utility in period t can now be written as:

$$U_A = w_A - \frac{1}{2}\theta e^2, \tag{3.1}$$

where $w_A \equiv a + db$ is the agent's total compensation. The agent's reservation utility equals $\overline{U_A} > 0$. Following Baker et al. (1994), the agent is assumed to play a trigger strategy that prescribes to always shirk if the supervisor once evaluated performance incorrectly ($d \neq V$).

The supervisor's utility in period t is given by:

$$U_S = w_S + \gamma U_A, \tag{3.2}$$

where w_S denotes the supervisor's compensation, and γ denotes the supervisor's altruism towards the agent. The supervisor's compensation is defined as $w_S \equiv \alpha + \eta z - \beta db$. The variable α is the supervisor's base salary. β ties the supervisor's compensation to his evaluation of the agent's performance. Finally, η is a bonus which is paid out if the verifiable signal about the agent's performance, z, is high ($z = 1$). Note that the supervisor's contract is also fully enforceable.

The supervisor may have standard selfish preferences, in which case $\gamma = \underline{\gamma} = 0$. Alternatively, the supervisor may be altruistic towards the agent, in which case $\gamma = \overline{\gamma} > 0$. It holds that $\overline{\gamma} < 1$, implying the altruistic supervisor always cares more for his own compensation than for the agent's utility. I assume that the supervisor's altruism is specific to the agent, and not a general trait towards all people. It follows that the supervisor's reservation utility, $\overline{U_S} > 0$, is independent of his type.[7] Further, the agent and the supervisor learn γ before they accept or reject the principal's job

[6] This is not a strong assumption, since the supervisor's report d can easily be made verifiable information (for example, d can be a written report). See also e.g. Prendergast and Topel (1996), Giebe and Gürtler (2012), and Thiele (2013), who make the same assumption.

[7] Of course, $\overline{U_S}$ may incorporate that the supervisor expects to have some degree of altruism towards another agent he may work with when taking his outside option.

offer.[8] Also, it is assumed that the principal does not observe γ.[9] Finally, for simplicity the probability that the supervisor is altruistic towards the agent equals $\Pr[\gamma = \overline{\gamma}] = \frac{1}{2}$, which is common knowledge.

The principal's profits are given by the agent's performance V, minus the agent's and the supervisor's compensation:

$$\Pi = V - w_S - w_A. \tag{3.3}$$

The principal's reservation utility $\overline{\Pi}$ is normalized to zero. To make the problem interesting, it is assumed that:

$$\frac{1}{2\theta} - \overline{U_A} - \overline{U_S} > 0. \tag{3.4}$$

Assumption (3.4) ensures that, if the agent exerts the first-best effort level and the agent and the supervisor don't earn a rent, hiring an agent and a supervisor is attractive.

The order of the game is as follows.

1) In some period t, the principal is matched to one agent and one supervisor.

2) The supervisor's type γ is randomly drawn, and observed by the agent and the supervisor.

3) The principal offers a take-it-or-leave-it contract to the agent and the supervisor. I impose that the supervisor's contract is uniform, that is, the principal cannot offer a menu of contracts to the supervisor.[10] All players observe the content of both contracts.

4) The agent and the supervisor decide whether or not to accept their contract. Without loss of generality, I assume that all players receive their reservation utility

[8] In light of the dynamic nature of the model, this is a realistic assumption. Alternatively, one could imagine that the agent and supervisor have been co-workers at the firm. During that period they formed some social relationship. The formal model then starts at the point when the principal offers one of the co-workers a promotion to the position of supervisor.

[9] As will become clear in the analysis, unless the supervisor's contract screens for one type, the principal cannot learn the supervisor's type over time either.

[10] A motivation for this assumption is the 'equal pay for equal work' principle. The assumption also rules out that, within a firm, some supervisors are imposed a penalty for giving high performance evaluations while others are not.

in period t unless both the agent and the supervisor accept their contract.

5) If both the agent and the supervisor accept their contract, the agent exerts effort, the supervisor gives a performance evaluation, and payoffs realize.

6) Period $t+1$ starts. If both the agent and the supervisor accepted their contract, the game is played again from the third stage. Otherwise, the game is played again from the first stage. In this case, the principal is matched to a new agent and a new supervisor.

3.4 Analysis

In this section I solve for the subgame perfect equilibrium of the game by using backward induction. Subsection 3.4.1 introduces the central agency problem, namely that the supervisor may want to bias performance evaluations. After that, I derive the agent's optimal effort choice and participation decision. In subsection 3.4.3 the optimal contracts for the supervisor and the agent are derived. At this stage, the principal's choice whether or not to screen for a supervisor type is still kept exogenous. The principal's screening decision is analyzed in subsection 3.4.4.

3.4.1 The supervisor's evaluation decision

After the agent has exerted effort and his performance V has realized, the supervisor decides to give a high or low performance rating to the agent. The evaluation of the agent's performance affects the supervisor's utility both in the current period t and in the future. The supervisor's marginal utility in period t from giving a high performance rating is $(\gamma - \beta)\, b$, where γ reflects the supervisor's altruistic utility from giving the agent a high performance evaluation and β reflects the supervisor's monetary costs from doing so. Depending on $\gamma \gtrless \beta$ and the realization of the agent's performance $V \in \{0, 1\}$, four cases can be distinguished. In two of these cases, the supervisor may have an incentive to bias performance evaluations. If $\gamma > \beta$ and $V = 0$, the supervisor may have an incentive to give an unwarranted high performance evaluation. If $\gamma < \beta$ and $V = 1$, the supervisor may have an incentive to give an unwarranted low performance evaluation. In both cases, the supervisor's

marginal benefit in period t from biasing the performance evaluation can be written as $|\gamma - \beta| b \geq 0$. In the remaining two cases, the marginal benefit in period t from biasing the performance evaluation equals $-|\gamma - \beta| b \leq 0$.

In the repeated game, the supervisor will only report performance truthfully in period t if the following holds: reporting performance truthfully yields rents in the future, the discounted value of which exceeds the current marginal benefit from biasing the performance evaluation. Formally, this requires:

$$\frac{\delta}{1-\delta}\left\{U_S^{tr} - \max\left[U_S^b; \overline{U_S}\right]\right\} \geq |\gamma - \beta| b, \text{ and} \tag{3.5}$$

$$\frac{\delta}{1-\delta}\left\{U_S^{tr} - \max\left[U_S^b; \overline{U_S}\right]\right\} \geq -|\gamma - \beta| b. \tag{3.6}$$

where U_S^{tr} is the supervisor's per-period utility from the job, given that he reports performance truthfully in period t, and U_S^b is the supervisor's per-period utility from the job, given that he biases the performance evaluation in period t. Note that, because he can always leave the firm, the supervisor earns at least his reservation utility after biasing the performance evaluation. Condition (3.5) therefore ensures that the supervisor enjoys a weakly positive rent from his job. In turn, this implies that the supervisor is always willing to accept the job. In addition to this, if his marginal benefit from biasing the performance evaluation equals $-|\gamma - \beta| b \leq 0$, the supervisor will always report performance truthfully. In other words, condition (3.6) is slack. In the remainder of this subsection, U_S^{tr} and U_S^b are derived explicitly.

Given that the supervisor reports performance truthfully, the agent will exert some optimal effort level e^* (to be derived in the next subsection). The probability that $V = 1$ equals e, and given that $E[\mu] = 1$, the expected probability that $z = 1$ also equals e. It follows that the supervisor's utility from the job when rating performance correctly is equal to:

$$U_S^{tr} = \alpha + e^*(\eta - \beta b) + \gamma U_A. \tag{3.7}$$

If the supervisor biased a performance evaluation, the agent will henceforth shirk. This implies that the signal z will always be low in the future. Even though the

agent shirks, if $\gamma > \beta$, the supervisor finds it nevertheless attractive to give a high evaluation report. In this case, the agent always earns the bonus b, implying he will stay with the firm.[11] However, if $\gamma < \beta$, the supervisor always claims performance was low, implying the agent only earns the base salary a. Depending on whether $a < \overline{U_A}$ or $a \geq \overline{U_A}$, the agent will leave or stay with the firm. In case the agent leaves the firm, the supervisor also loses his job by assumption. Hence, the supervisor's per-period utility from the job after biasing the performance evaluation is equal to:

$$U_S^b = \begin{cases} \alpha + \gamma a + (\gamma - \beta) b & \text{if } \gamma \geq \beta \\ \alpha + \gamma a & \text{if } \gamma < \beta \text{ and } a \geq \overline{U_A} \\ \overline{U_S} & \text{if } \gamma < \beta \text{ and } a < \overline{U_A} \end{cases}. \tag{3.8}$$

Finally, the supervisor already enjoys either U_S^{tr} or U_S^b in the first period he is employed. The reason is that the agent knows whether or not the supervisor will evaluate performance truthfully. This follows from the assumptions that the agent observes the supervisor's type and contract.

3.4.2 The agent's effort and participation decision

Given that the supervisor reports performance truthfully, the agent's optimal effort choice is:

$$e = \frac{b}{\theta}. \tag{3.9}$$

Note that, since the first best effort-level equals $e^{FB} = \frac{1}{\theta}$, the first-best level of the bonus equals $b^{FB} = 1$. If the supervisor does not report performance truthfully, the agent will exert zero effort. The reason is that the supervisor will give either

[11] To see this, suppose that the supervisor will report performance truthfully, implying the agent will optimally exert some effort level e^* (to be derived in the next section). To make the job attractive to the agent, the principal must then pay the agent a base salary at least equal to:

$$a^* \geq \overline{U_A} + \frac{1}{2}\theta \left(e^*\right)^2 - e^* b.$$

Next, suppose that, after both players have accepted the job, the supervisor chooses to bias the performance evaluation upward. The agent then earns the bonus irrespective of his performance, and so it becomes optimal for him to shirk. Given the minimum base salary derived above, the agent's utility from the job now becomes $a^* + b \geq \overline{U_A} + \frac{1}{2}\theta \left(e^*\right)^2 + (1 - e^*) b$. Since $e \leq 1$ by assumption, this is always greater than the agent's reservation utility $\overline{U_A}$.

a high or low performance evaluation, independent of the agent's performance. In that case, exerting effort yields the agent no marginal benefit.

Any contract offered to the agent should satisfy the agent's participation constraint. Given the agent's effort choice, and assuming the supervisor reports performance truthfully, this constraint reads:

$$a + \frac{1}{2}\frac{b^2}{\theta} \geq \overline{U_A}. \tag{3.10}$$

3.4.3 Optimal contracts

In this subsection the optimal contracts for the agent and supervisor are derived. The principal's screening decision is still kept exogenous. It is analyzed in the following subsection. I start this subsection by establishing the following intermediary result. Suppose that the principal wants to attract some supervisor type and have him report performance truthfully. Then, depending on the supervisor's contract, it may be that $U_S^b > \overline{U_S}$ or $U_S^b \leq \overline{U_S}$ (see (3.8)). However, the principal is better off the lower U_S^b is. The intuition is that, the lower the supervisor's utility after biasing a performance evaluation, the less total compensation he requires to be willing to report performance truthfully. Since the supervisor can always take his outside option, a lower bound on U_S^b is given by $\overline{U_S}$. It follows that the principal strictly prefers $U_S^b \leq \overline{U_S}$ to $U_S^b > \overline{U_S}$.

The next thing to note is that the principal can set the supervisor's base salary α as low as necessary to ensure that $U_S^b \leq \overline{U_S}$ holds (see (3.8)). Moreover, doing so is costless. This holds because the base salary α and the bonus η are substitutes in the supervisor's compensation, given that he reports performance truthfully (see (3.7)). A low base salary can thus be compensated for by a higher value of η. Lemma 3.1 follows:

Lemma 3.1 *The principal sets the supervisor's base salary α sufficiently low, such that $U_S^b \leq \overline{U_S}$ holds for each supervisor type the principal attracts.*

Before we turn to the optimal contracts, note that Lemma 3.1 implies that the

supervisor will report performance truthfully if:

$$\frac{\delta}{1-\delta}\left[\alpha + e^*\left(\eta - \beta b\right) + \gamma U_A - \overline{U_S}\right] \geq \left|\gamma - \beta\right| b. \tag{3.11}$$

Attracting one supervisor type (screening)

If screening for one supervisor type is optimal, the principal offers the following contracts to the agent and supervisor.

Proposition 3.1 *Suppose screening for one supervisor type is optimal. The supervisor's optimal contract only attracts the altruistic type. The supervisor's base salary* α^* *satisfies Lemma 3.1, and it holds that:*

$$\beta^* = \overline{\gamma} > 0, \text{ and } \eta^* \geq \frac{1}{2}\overline{\gamma} > 0.$$

The supervisor's total compensation equals $w_S^* = \overline{U_S} - \overline{\gamma}\overline{U_A}$.

The agent's optimal contract is given by:

$$a^* = \overline{U_A} - \frac{1}{2\theta} < \overline{U_A} \text{ and } b^* = b^{FB} = 1.$$

The agent earns his reservation utility from the job, $U_A = \overline{U_A}$.

Proof. Appendix 3.A. ■

The optimal contracts can be explained as follows. First, screening reveals the supervisor's type which is valuable information to the principal. The reason is that it enables her to make the supervisor indifferent to his evaluation report. This is achieved by equating β to the supervisor's type γ. The benefit is that the supervisor will report performance truthfully without requiring costly rents to do so (one can easily see this in condition (3.5)). Moreover, since the supervisor is indifferent to his evaluation report for any value of b, it is optimal to pay the agent the first-best level of the bonus, $b^* = b^{FB} = 1$. Note that because the optimal screening contract attracts only the altruistic type, the contract contains a monetary cost for reporting high performance ($\beta^* = \overline{\gamma} > 0$).

Second, the reason that the supervisor's contract screens for the altruistic type is as follows. As explained above, given that $\beta = \gamma$, the supervisor requires the lowest possible compensation level to report performance truthfully, which equals $w_S = \overline{U_S} - \gamma U_A$. Importantly, since it holds that $\overline{\gamma} > \underline{\gamma} = 0$, the altruistic type requires a lower compensation level than the selfish type. The intuition is that in contrast to the selfish type, the altruistic type enjoys working with the agent. Next, since both types are equally likely to occur, the supervisor is equally likely to accept his contract independent of the type screened for. It follows that the principal's expected profits from screening are highest if she screens for the altruistic supervisor.[12]

Third, the supervisor's compensation depends positively on the signal z via the bonus $\eta^* > 0$. The intuition is that the supervisor is optimally paid a low base salary α^*, as implied by Lemma 3.1. However, it must still hold that reporting performance truthfully is sufficiently attractive for the altruistic type. This is ensured by the bonus η^*.

Fourth, to see why the supervisor's contract only attracts the altruistic type, note that the wage differential explained above implies that reporting performance truthfully yields the selfish supervisor less than his reservation utility. Hence, the selfish type may only be hired given that he biases the agent's performance evaluation. If the selfish supervisor biases the performance evaluation, he will bias the evaluation downward. The reason is that giving a high performance evaluation imposes a monetary cost upon the supervisor ($\beta^* = \overline{\gamma} > 0$). However, expecting this, the agent will not accept his contract in the first place. This is true because the agent will only earn the base salary from the job, which is below his reservation utility. If the agent rejects his contract, I assume that the principal does not hire the supervisor

[12] If, contrary to the assumption made, the selfish type is sufficiently more abundant than the altruistic type, the principal may prefer to screen for the selfish type. It is possible to show that such a screening contract exists if it holds that $\frac{\delta}{1-\delta}\overline{U_A} < 1$. If this contract is accepted by the supervisor, the principal's profits are the same as in Proposition 3.1, minus the compensating wage differential the supervisor's altruism gives rise to. The possibility that screening for the selfish type dominates screening for the altruistic type only affects the first result of Proposition 3.3: for low values of the discount factor, it may be optimal to screen for the selfish type rather than for the altruistic type. However, the intuition behind the optimality of screening per se remains the same, namely that screening enables the principal to incentivize the supervisor at lower costs.

either. It follows that the screening contract only attracts the altruistic supervisor type.

Last, the principal pays the agent such a low base salary that the agent earns only his reservation utility from the job. This is optimal even though raising the agent's utility decreases the altruistic supervisor's compensation. The reason is that the supervisor always cares more for his own utility than the agent's utility ($\overline{\gamma} < 1$). It is therefore not possible for the principal to gain from paying the agent a higher compensation and cut the supervisor's compensation more than proportionally.

By way of summary, if the principal screens for the altruistic supervisor, the supervisor nor the agent earn a rent, and the agent exerts the first-best effort level. Therefore, screening yields the principal first-best profits, conditional on the supervisor and the agent accepting their contract.

Attracting both supervisor types (pooling)

If the principal optimally attracts both supervisor types, she offers the following contracts to the agent and supervisor.

Proposition 3.2 *Suppose the principal optimally attracts both supervisor types. Denote by* δ^L, δ^M, *and* δ^H *threshold values of the discount factor, where it holds that* $0 < \delta^L < \delta^M < \delta^H < 1$. *There are three different cases:*

Case I applies for $\delta \in [\delta^H, 1)$, *in which it holds that:*

$$\begin{aligned} \beta_I^{**} &= 0, \\ w_{S,I}^{**} &= \overline{U_S}, \text{ and} \\ b_I^{**} &= 1. \end{aligned}$$

Case II applies for $\delta \in \left(0, \delta^L\right]$ *and* $\delta \in \left[\delta^M, \delta^H\right)$. *The only difference with case I is that:*

$$b_{II}^{**} = \frac{\delta}{1-\delta}\overline{U_A} < 1.$$

Case III applies for $\delta \in \left(\delta^L, \delta^M\right)$*, in which it holds that:*

$$\begin{aligned} \beta^{**}_{III} &= \frac{\overline{\gamma}}{2}\left(1 - \frac{\delta}{1-\delta}\frac{\overline{U_A}}{b^{**}_{III}}\right), \\ w^{**}_{S,III} &= \overline{U_S} - \frac{\overline{\gamma}}{2}\overline{U_A} + \frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta}b^{**}_{III}, \text{ and} \\ b^{**}_{III} &= 1 - \theta\frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta} < 1. \end{aligned}$$

It holds that $b^{**}_I > b^{**}_{II} > b^{**}_{III} > 0$*,* $w^{**}_{S,I} = w^{**}_{S,II} < w^{**}_{S,III}$*, and* $\beta^{**}_{III} > 0$*. Furthermore, for all cases it holds that 1) the agent's base salary* a^{**} *is set such that, given his effort choice, the bonus* b^{**}*, and the supervisor's contract, the agent earns his reservation utility from the job (*$U_A = \overline{U_A}$*), 2) the agent's base salary is below the agent's reservation utility (*$a^{**} < \overline{U_A}$*), 3) the supervisor's base salary* α^{**} *satisfies Lemma 3.1, and 4) the supervisor receives a bonus conditional on the signal about the agent's performance being high (*$\eta^{**} > 0$*).*

Proof. Appendix 3.A. ■

All cases from Proposition 3.2 share the property that the bonus $\eta^{**} > 0$. The intuition is identical to the one given in the previous subsection. The supervisor is paid a low base salary to make biasing performance ratings unattractive (Lemma 3.1). The bonus $\eta^{**} > 0$ compensates the supervisor for the low base salary, given that he reports performance truthfully. Also, the agent always earns his reservation utility from the job. This is optimal, simply because it minimizes the agent's total compensation level. In the remainder of this subsection, the distinguishing properties of each case from Proposition 3.2 are discussed in detail.

Case I. The most remarkable property of case I is that it sets $\beta^{**}_I = 0$. The supervisor's compensation is thus independent of his evaluation of the agent's performance. By the intuition from Proposition 3.1, this is the most efficient way to induce truthful ratings from the selfish type. However, by the same logic, $\beta^{**}_I = 0$ is very inefficient when inducing truthful ratings from the altruistic type. To see why $\beta^{**}_I = 0$ is nevertheless optimal, recall first that the supervisor will report

performance truthfully if:

$$\frac{\delta}{1-\delta}\left(w_S + \gamma U_A - \overline{U_S}\right) \geq |\gamma - \beta|\, b. \tag{3.12}$$

Second, given $\beta = 0$, this condition is only satisfied for the selfish type ($\gamma = 0$) if it holds that $w_S \geq \overline{U_S}$. Note that the supervisor cannot be paid a lower compensation level, since otherwise the selfish type will not accept the job. Third, given $\beta = 0$ and $w_S = \overline{U_S}$, condition (3.12) for the altruistic type reduces to:

$$\frac{\delta}{1-\delta} U_A \geq b. \tag{3.13}$$

Hence, if the above condition holds, the altruistic type will report performance truthfully even if there is no monetary penalty attached to reporting high performance, and the supervisor is paid the lowest possible compensation level.

The intuition behind this result is that the altruistic supervisor enjoys working with the agent, as measured by γU_A. As long as he reports performance truthfully, this utility cannot be extracted from the altruistic type because doing so would violate the selfish type's participation constraint. However, after biasing a performance evaluation, the optimal contract ensures that the supervisor's utility equals at most his reservation utility (Lemma 3.1). The altruistic supervisor thus enjoys a rent as long as he rates performance correctly. The implication is that altruism not only provides an incentive to bias performance evaluations, it also provides a motivation to report performance truthfully. Moreover, if the supervisor is sufficiently patient, as measured by δ, the latter effect outweighs the former. If this is the case, altruism provides a net incentive to report performance truthfully, rather than to bias performance evaluations. This is the first key result of the chapter:

Corollary 3.1 *If condition (3.13) holds, supervisor's altruism provides a net incentive to report performance truthfully, rather than to bias performance evaluations.*

Next, in case I, the supervisor is assumed to be so patient ($\delta \geq \delta^H$) that the altruistic type prefers to report performance truthfully, even if $b = 1$ and $U_A = \overline{U_A}$. It follows that both supervisor types report performance truthfully, the agent is

provided first-best incentives, and the agent and the supervisor are paid the lowest possible compensation level. There is no set of contracts that could yield higher profits. Finally, note that for reasonable values of the discount factor, say $\delta \geq 0.8$, condition (3.13) is satisfied even if the agent's first-best bonus is a factor four higher than the agent's outside option utility. Therefore, case I is likely to apply in real workplace settings.

Case II. This case is almost identical to case I, except that it applies for lower values of the supervisor's patience ($\delta < \delta^H$). The implication is that, if $b = 1$ and $U_A = \overline{U_A}$, supervisor altruism no longer provides a net incentive to report performance truthfully, but rather to bias evaluations. In case II, the principal deals with this problem in a simple way: she reduces the agent's bonus till the point where the supervisor's altruism again provides a net incentive to report performance truthfully. Hence, $b_{II}^{**} = \frac{\delta}{1-\delta}\overline{U_A} < 1$.

One may note that an alternative way to solve this problem is to raise the agent's utility above $\overline{U_A}$. This is never optimal, however. The reason is that the supervisor always cares more for his own compensation than for the agent's utility ($\overline{\gamma} < 1$). It is therefore more efficient to increase the supervisor's rents directly by raising his own compensation, rather than indirectly by raising the agent's utility. This brings us to the final case.

Case III. As does case II, case III applies for relatively low values of the discount factor. However, in contrast to case II, the principal does not design the agent's and supervisor's contract such that the supervisor's altruism provides a net incentive to report performance truthfully. Instead, she simply raises the supervisor's compensation to ensure that reporting truthfully yields sufficient rents for the altruistic type. This response gives rise to two trade-offs.

First, the principal optimally lowers the agent's bonus below the first-best level. The reason is that the supervisor would enjoy a lower marginal benefit in period t from biasing the performance evaluation. The benefit to the principal is that the supervisor now requires lower rents from reporting truthfully, which in turn lowers his compensation.

Second, the principal imposes a penalty on the supervisor for giving a high performance evaluation, that is, $\beta_{III}^{**} > 0$. To see why this is optimal, note that raising the supervisor's compensation not only increases the altruistic type's rents from reporting truthfully, but also the selfish type's. The principal can therefore impose a penalty upon the supervisor for giving a high performance evaluation, without changing the selfish type's decision to report truthfully. The benefit of doing so is that the altruistic type enjoys a lower marginal benefit in period t from biasing performance evaluations. This, in turn, lowers the supervisor's compensation. Whether case III yields higher profits than case II depends on the value of the discount factor. In the proof of the Proposition, the threshold values of the discount factor are explicitly derived.

Summarizing this subsection, offering the pooling contract to the supervisor yields first-best profits if the discount factor is high (case I). For lower values of the discount factor the principal's profits are below first-best because the agent's incentives are reduced (case II), or because the agent's incentives are reduced and the supervisor earns a rent (case III).

3.4.4 Optimal screening

In this subsection, the principal's decision whether or not to offer the screening contract to the supervisor is analyzed. Denote by $E[\Pi]$ the principal's per-period profits given that the agent and the supervisor accepted their contract, which will be defined later on. Next, if the principal offers the pooling contract to the supervisor, there is a probability $\varphi = 0$ that the supervisor will reject his contract, whereas if the principal offers the screening contract there is a probability of $\varphi = \frac{1}{2}$ that the supervisor rejects his contract. If the supervisor rejects his contract, all players by assumption earn their outside option in period t. In that case the principal is matched to a new agent and a new supervisor in period $t+1$. The principal's expected lifetime utility from being matched to one agent and one supervisor can now be written as:

$$E\left[\Pi^{L}\right] = \varphi\left\{0 + \delta E\left[\Pi^{L}\right]\right\} + (1-\varphi)\frac{1}{1-\delta}E[\Pi], \tag{3.14}$$

which is equivalent to:

$$E\left[\Pi^{L}\right] = \frac{1-\varphi}{(1-\delta)(1-\delta\varphi)} E\left[\Pi\right]. \tag{3.15}$$

Let's now define the value of $E[\Pi]$. Using the results from Proposition 3.1, if the principal screens for the altruistic supervisor, $E[\Pi]$ equals:

$$E\left[\Pi^{*}\right] = \frac{1}{2\theta} - (1-\overline{\gamma})\,\overline{U_A} - \overline{U_S} > 0, \tag{3.16}$$

where the sign follows from assumption (3.4) and $\overline{\gamma} > 0$. If the principal attracts both supervisor types, the value of $E[\Pi]$ depends on which case from Proposition 3.2 applies. For each case, the principal's profits are given by:

$$E\left[\Pi_{I}^{**}\right] = \frac{1}{2\theta} - \overline{U_A} - \overline{U_S} > 0, \tag{3.17}$$

$$E\left[\Pi_{II}^{**}\right] = \frac{b_{II}^{**}}{\theta} - \frac{1}{2}\frac{\left(b_{II}^{**}\right)^{2}}{\theta} - \overline{U_A} - \overline{U_S}, \text{ and} \tag{3.18}$$

$$E\left[\Pi_{III}^{**}\right] = \frac{b_{III}^{**}}{\theta} - \frac{1}{2}\frac{\left(b_{III}^{**}\right)^{2}}{\theta} - \frac{1-\delta}{\delta}\frac{\overline{\gamma}}{2} b_{III}^{**} - \left(1 - \frac{\overline{\gamma}}{2}\right)\overline{U_A} - \overline{U_S}, \tag{3.19}$$

respectively. It is now possible to derive when screening is optimal by comparing the expected lifetime utility from screening to the expected lifetime utility from pooling.

By Proposition 3.2, if the principal offers the pooling contract to the supervisor and $\delta \geq \delta^{H}$, the contracts from case I are optimal. It follows that, conditional on $\delta \geq \delta^{H}$, screening is optimal if:

$$\frac{1-\frac{1}{2}}{(1-\delta)\left(1-\frac{1}{2}\delta\right)} E\left[\Pi^{*}\right] \geq \frac{1}{1-\delta} E\left[\Pi_{I}^{**}\right]. \tag{3.20}$$

Rewriting this condition yields that screening is optimal if the discount factor is sufficiently high:

$$\delta \geq 2 - \frac{E\left[\Pi^{*}\right]}{E\left[\Pi_{I}^{**}\right]} \tag{3.21}$$

Note that the threshold value of the discount factor is smaller than 1 since $E[\Pi^{*}] > E[\Pi_{I}^{**}]$. The intuition behind this result is as follows. If the principal screens for the altruistic type, conditional on the supervisor accepting his contract, she enjoys

the compensating wage differential that altruism gives rise to. I call this motive the 'wage differential motive' for screening. As revealed by condition (3.21), the wage differential motive only makes screening attractive if the principal is sufficiently patient. The reason is that screening implies that the supervisor may reject his job a number of times. In expectation, the wage differential is therefore only enjoyed somewhere in the future.

Next, by Proposition 3.2, if the principal offers the pooling contract to the supervisor and $\delta < \delta^H$, the contracts from case II or case III are optimal. Unfortunately, the equivalent of condition (3.20) for cases II and III is difficult to solve explicitly for the discount factor. However, by the following argument it can still be shown that screening is optimal for low values of the discount factor. First, recall from Proposition 3.2 that, if the principal attracts both supervisor types and $\delta \in \left(0, \delta^L\right]$, case II applies. Hence, the principal's profits from attracting both supervisor types are given by $E\left[\Pi_{II}^{**}\right]$. Second, $E\left[\Pi_{II}^{**}\right]$ is lower for lower values of the discount factor.[13] The reason is that the agent's bonus b_{II}^{**} is lower for lower values of the discount factor. As explained in the previous subsection, this is done to ensure that the altruistic supervisor prefers to report performance truthfully. Third, in the limit where the discount factor is zero, the agent's bonus is optimally equal to zero. It follows that $E\left[\Pi_{II}^{**}\right]$ is strictly negative. However, the principal's profits from screening are always strictly positive. Therefore, there must be a range of low discount factors for which screening is optimal. The intuition behind this result is that screening allows the principal to give the agent stronger incentives. I call this the 'incentive motive' for screening. The final Proposition follows:

Proposition 3.3 *There exist two ranges of discount factors, $\delta \in (0, \underline{\delta}]$ and $\delta \in [\overline{\delta}, 1)$ where $\underline{\delta} > 0$ and $\overline{\delta} < 1$, for which the principal optimally screens for altruistic supervisors.*

Finally, note that a similar limit argument as given above for case II cannot be

[13] Taking derivatives yields:

$$\frac{\partial E\left[\Pi_{II}^{**}\right]}{\partial \delta} = \frac{1 - b_{II}^{**}}{\theta\left(1-\delta\right)^2} > 0,$$

where the sign follows from the fact that $b_{II}^{**} < 1$.

established for case III. The simple reason is that case III only applies for intermediate values of the discount factor (see Proposition 3.2). However, as in case II, it holds that the principal's profits from attracting both supervisor types are lower, the lower the value of the discount factor.[14] The reason is that, in order to reduce the supervisor's compensation, the agent is optimally paid a lower bonus if the discount factor is lower. Hence, also in case III the incentive motive may imply that screening is optimal for relatively low values of the discount factor. In addition to this, note that in both cases II and III, the wage differential motive may imply that screening is optimal for relatively high values of the discount factor. However, due to computational limitations, it cannot be established when these motives are sufficiently strong such that screening is optimal.

3.5 Concluding remarks

This chapter studied how firms can optimally deal with the possibility of leniency bias through designing contracts for employees and potentially altruistic supervisors. Two key results emerge from this analysis. First, the optimal contracts sometimes ensure that supervisor's altruism provides a net incentive to report performance truthfully, rather than to bias performance evaluations. Second, if the discount factor is relatively low or high, the principal optimally screens for the altruistic supervisor.

The present analysis assumes that the principal can only offer a uniform contract to the supervisor, which seems to be a realistic assumption in many workplace settings. Nevertheless, if the principal has the possibility to offer a menu of contracts to the supervisor it could be optimal for her to exploit this possibility, and hence the results found here may be affected. I leave this potentially interesting case for further research.

Another avenue for further research is to allow for the possibility of collusion.

[14] Taking derivatives yields:

$$\frac{\partial E\left[\Pi_{III}^{**}\right]}{\partial \delta} = \frac{\overline{\gamma}}{4\delta^3}\left[2\delta - (1-\delta)\,\overline{\gamma}\theta\right] > 0,$$

where the sign follows from the fact that $b_{III}^{**} > 0$.

A detailed analysis of collusion proof contracts in an infinitely repeated game is provided by Thiele (2013). Interestingly, the analysis by Thiele (2013) and the one presented here display close similarities. Specifically, an altruistic supervisor can be interpreted as a supervisor who is prone to colluding with the agent against the principal. Both Thiele (2013) and the present analysis reveal that, to eliminate bias in performance evaluation, the supervisor may require a rent, and the agent's incentives may be optimally reduced. It would be interesting to study whether and how social relationships and the possibility of collusion interact in the workplace.

3.A Proofs

This Appendix contains the proofs of the results described in Propositions 3.1 and 3.2. To facilitate the reading of the proofs, I start by listing the constraints the principal faces when designing the agent's and supervisor's contracts.

First, given that the supervisor reports performance truthfully, the agent's incentive compatibility constraint is:

$$e = \frac{b}{\theta}, \tag{3.22}$$

and, second, the agent's participation constraint is:

$$U_A = a + \frac{1}{2}\frac{b^2}{\theta} \geq \overline{U_A}. \tag{3.23}$$

Last, the altruistic and selfish supervisor types' truthtelling constraints are given by:

$$\frac{\delta}{1-\delta}\left[\alpha + \frac{b}{\theta}(\eta - \beta b) + \overline{\gamma}U_A - \overline{U_S}\right] \geq |\overline{\gamma} - \beta|\, b, \text{ and} \tag{3.24}$$

$$\frac{\delta}{1-\delta}\left[\alpha + \frac{b}{\theta}(\eta - \beta b) - \overline{U_S}\right] \geq |0 - \beta|\, b, \tag{3.25}$$

respectively. Note that (3.24) and (3.25) ensure that reporting performance truthfully yields the supervisor at least his reservation utility. Finally, recall that α must satisfy Lemma 3.1. That is, α must be set sufficiently low such that $U_S^b \leq \overline{U_S}$ holds for each supervisor type the principal attracts, where U_S^b is given by (3.8).

Proof of Proposition 3.1

In this proof the optimal contracts are derived, given that the principal optimally screens for one supervisor type. The proof relies on the assumption that the principal optimally screens for the altruistic supervisor type. In the main text it is proved that, and explained why this assumption holds.

The principal's problem is to design the contracts that maximize her expected profits subject to the agent's incentive compatibility constraint (3.22), the agent's participation constraint (3.23), and the altruistic supervisor type's truthtelling constraint (3.24). Moreover, the supervisor's contract must not attract the selfish super-

visor type. I start by deriving the optimal contracts assuming the latter condition is satisfied. After that, it is checked whether this condition holds, given the optimal contract for the supervisor. The principal's problem can now be written as:

$$\max_{a,b,\beta,\eta} \quad E\left[\Pi\right] = \frac{b}{\theta} - \left(a + \frac{b^2}{\theta}\right) - \left[\alpha + \frac{b}{\theta}\left(\eta - \beta b\right)\right] \text{ s.t. (3.23) and (3.24).}$$

The first-order conditions to the problem are:

$$\begin{aligned}
\frac{\partial E\left[\Pi\right]}{\partial a} &= -1 + \lambda_1 + \lambda_2 \frac{\delta}{1-\delta}\overline{\gamma} = 0, && (3.26)\\
\frac{\partial E\left[\Pi\right]}{\partial b} &= \frac{1}{\theta} - 2\frac{b}{\theta} - \left(\frac{1}{\theta}\eta - 2\frac{b}{\theta}\beta\right) + \lambda_1 \frac{b}{\theta} + && (3.27)\\
&\quad \lambda_2\left[\frac{\delta}{1-\delta}\left(\frac{1}{\theta}\eta - 2\frac{b}{\theta}\beta + \overline{\gamma}\frac{b}{\theta}\right) - \left|\overline{\gamma} - \beta\right|\right] = 0,\\
\frac{\partial E\left[\Pi\right]}{\partial \beta} &= \frac{b^2}{\theta} - \lambda_2\left[\frac{\delta}{1-\delta}\frac{b^2}{\theta} + \frac{\beta - \overline{\gamma}}{\left|\overline{\gamma} - \beta\right|}b\right] = 0, \text{ and} && (3.28)\\
\frac{\partial E\left[\Pi\right]}{\partial \eta} &= -\frac{b}{\theta} + \lambda_2 \frac{\delta}{1-\delta}\frac{b}{\theta} = 0, && (3.29)
\end{aligned}$$

where λ_1 and λ_2 are the Lagrange multipliers on the agent's participation constraint (3.23) and the altruistic supervisor's truthtelling constraint (3.24), respectively. Solving the system of first-order conditions yields $\lambda_1 = 1 - \overline{\gamma} > 0$, $\lambda_2 = \frac{1-\delta}{\delta} > 0$, $\beta^* = \overline{\gamma}$, and $b^* = 1$. From $\lambda_1 > 0$ it follows that the agent's participation constraint (3.23) binds. This implies the agent's base salary equals $a^* = \overline{U_A} - \frac{1}{2\theta} < 0$, where the sign follows from assumption (3.4). Also, the agent's utility from the job equals his reservation utility, $U_A = \overline{U_A}$. From $\lambda_2 > 0$ it follows that the altruistic supervisor's truthtelling constraint (3.24) binds. Solving (3.24), given the solutions found so far, yields $\eta^* = \overline{\gamma} + \theta\left(\overline{U_S} - \alpha - \overline{\gamma}\overline{U_A}\right)$, where the supervisor's base salary α must satisfy Lemma 3.1. Next, α satisfies Lemma 3.1 if $U_S^b \leq \overline{U_S}$ holds for the altruistic supervisor. Taking into account that $\beta^* = \overline{\gamma}$, equation (3.8) yields this is the case if $\alpha \leq \overline{U_S} - \overline{\gamma}a^*$. It follows that $\eta^* \geq \frac{1}{2}\overline{\gamma} > 0$. Further, for any α that satisfies Lemma 3.1, the supervisor's total compensation level can now be computed to be equal to $w_S^* = \overline{U_S} - \overline{\gamma}\overline{U_A}$.

It must still be checked whether the supervisor's contract indeed only attracts the altruistic type. First note that, given the solutions found so far, the selfish type's

truthtelling constraint (3.25) reduces to:

$$-\frac{\delta}{1-\delta}\overline{U_A} \geq 1, \tag{3.30}$$

which can never hold. The selfish type will therefore give biased performance evaluations if he accepts the job. Next, since it holds that the supervisor's monetary cost for reporting high performance is positive, $\beta^* = \overline{\gamma} > 0$, the selfish supervisor will always give a low performance evaluation. Therefore, the agent only earns his base salary from his job, which is below his reservation utility, $a^* < \overline{U_A}$. It follows that the agent will not accept the job if the supervisor is selfish. Finally, by assumption, in this case the selfish supervisor is not hired either.

Proof of Proposition 3.2

In this proof the optimal contracts are derived, given that the principal optimally attracts both supervisor types. I start by establishing the following two intermediate results.

First, it can never be optimal to attract both supervisor types, but have only one of them report performance truthfully. The reason is that this strategy is always dominated by the strategy to screen for one supervisor type. To see this, note that if the principal screens, she can enjoy two possible profit levels both of which occur with a probability of $\frac{1}{2}$. If the supervisor accepts his job, profits equal the first-best level (Proposition 3.1). If the supervisor rejects his job, by assumption the principal enjoys her reservation utility $\overline{\Pi} = 0$. In case both supervisor types are attracted but only one of them reports performance truthfully, there are also two possible profit levels that each occur with a probability of $\frac{1}{2}$. If the supervisor reports performance truthfully, the principal enjoys at most the first-best profit level. However, if the supervisor biases performance evaluations, profits are strictly negative. The reason is that the agent will exert zero effort in this case, while hiring the agent and supervisor is costly. Hence, it can never be optimal to attract both supervisor types, but incentivize only one of the types to report performance truthfully. It follows that both supervisor types' truthtelling constraints (3.24 and 3.25) must be satisfied in equilibrium.

Second, it must always hold that $\beta \in [0, \overline{\gamma}]$. The reason is that if $\beta < 0$ or $\beta > \overline{\gamma}$, the principal could increase or decrease β and thereby lower both types' marginal benefit in period t from rating performance incorrectly. This would be strictly profit-increasing, since then both types require lower rents to be willing to report performance truthfully. Because it turns out that only the lower constraint on β may bind in equilibrium, the following constraint is added to the principal's maximization problem:

$$\beta \geq 0. \tag{3.31}$$

The principal's problem can now be written as:

$$\max_{a,b,\beta,\eta} \quad E[\Pi] = \frac{b}{\theta} - \left(a + \frac{b^2}{\theta}\right) - \left[\alpha + \frac{b}{\theta}(\eta - \beta b)\right] \text{ s.t. (3.23), (3.24), (3.25), and (3.31).}$$

The first-order conditions to the problem are:

$$\frac{\partial E[\Pi]}{\partial a} = -1 + \lambda_1 + \lambda_2 \frac{\delta}{1-\delta}\overline{\gamma} = 0, \tag{3.32}$$

$$\begin{aligned}\frac{\partial E[\Pi]}{\partial b} = {} & \frac{1}{\theta} - 2\frac{b}{\theta} - \left(\frac{1}{\theta}\eta - 2\frac{b}{\theta}\beta\right) + \lambda_1 \frac{b}{\theta} + \\ & \lambda_2 \left[\frac{\delta}{1-\delta}\left(\frac{1}{\theta}\eta - 2\frac{b}{\theta}\beta + \overline{\gamma}\frac{b}{\theta}\right) - |\overline{\gamma} - \beta|\right] + \\ & \lambda_3 \left[\frac{\delta}{1-\delta}\left(\frac{1}{\theta}\eta - 2\frac{b}{\theta}\beta\right) - |0 - \beta|\right] = 0,\end{aligned} \tag{3.33}$$

$$\begin{aligned}\frac{\partial E[\Pi]}{\partial \beta} = {} & \frac{b^2}{\theta} - \lambda_2 \left[\frac{\delta}{1-\delta}\frac{b^2}{\theta} + \frac{\beta - \overline{\gamma}}{|\overline{\gamma} - \beta|}b\right] - \\ & \lambda_3 \left[\frac{\delta}{1-\delta}\frac{b^2}{\theta} + \frac{\beta - 0}{|0 - \beta|}b\right] + \lambda_4 = 0, \text{ and}\end{aligned} \tag{3.34}$$

$$\frac{\partial E[\Pi]}{\partial \eta} = -\frac{b}{\theta} + \lambda_2 \frac{\delta}{1-\delta}\frac{b}{\theta} + \lambda_3 \frac{\delta}{1-\delta}\frac{b}{\theta} = 0, \tag{3.35}$$

where λ_1, λ_2, λ_3, and λ_4 are the Lagrange multipliers on (3.23), (3.24), (3.25), and (3.31), respectively. There are three qualitatively different sets of solutions to the system of first-order conditions. Which one applies will be shown to depend on the value of the discount factor δ. I start by solving the case that has been labelled case III in the main text.

Case III. Let's first assume that the constraint on β does not bind, that is, $\beta > 0$

and $\lambda_4 = 0$. Solving (3.34) and (3.35) then yields $\lambda_2 = \lambda_3 = \frac{1}{2}\frac{1-\delta}{\delta} > 0$, implying that both supervisor types' truthtelling constraints (3.24 and 3.25) bind in equilibrium. Solving these constraints for β and η yields:

$$\beta^{**}_{III} = \frac{\overline{\gamma}}{2}\left(1 - \frac{\delta}{1-\delta}\frac{U_A}{b_{III}}\right), \text{ and} \tag{3.36}$$

$$\eta^{**}_{III} = \frac{1-\delta}{\delta}\left(\overline{\gamma} - \beta^{**}_{III}\right)\theta - \frac{\theta}{b_{III}}\left[\alpha_{III} - \frac{b^2_{III}}{\theta}\beta^{**}_{III} + \overline{\gamma}U_A - \overline{U_S}\right], \tag{3.37}$$

where the supervisor's base salary α_{III} must satisfy Lemma 3.1. Solving (3.32) yields $\lambda_1 = 1 - \frac{1}{2}\overline{\gamma} > 0$. From this it follows that the agent's participation constraint (3.23) binds, implying that $a^{**}_{III} = \overline{U_A} - \frac{1}{2}\frac{\left(b^{**}_{III}\right)^2}{\theta}$ and $U_A = \overline{U_A}$. Given the solutions found so far, solving (3.33) yields:

$$b^{**}_{III} = 1 - \theta\frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta} < 1. \tag{3.38}$$

One can now compute the supervisor's total compensation level, which is equal to:

$$w^{**}_{S,III} = \overline{U_S} - \frac{\overline{\gamma}}{2}\overline{U_A} + \frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta}b^{**}_{III}. \tag{3.39}$$

Next, recall that Lemma 3.1 states that it must hold for all supervisor types attracted that the utility from biasing the performance evaluation is weakly below the supervisor's reservation utility, that is, $U^b_S \leq \overline{U_S}$. Taking into account that $\overline{\gamma} > \beta^{**}_{III} > 0$ and $a^{**}_{III} < \overline{U_A}$, equation (3.8) reveals this is the case for the altruistic type if it holds that the supervisor's base salary satisfies the following condition: $\alpha_{III} \leq \overline{U_S} - \overline{\gamma}a^{**}_{III} - \left(\overline{\gamma} - \beta^{**}_{III}\right)b^{**}_{III}$. In case the supervisor is selfish, it always holds that $U^b_S \leq \overline{U_S}$. The reason is that the agent will not accept the job given that the selfish type biases the performance evaluation, as the selfish type biases the performance evaluation downwards. By assumption, in this case the selfish supervisor is not hired either. Further, the condition imposed on the base salary α_{III} by Lemma 3.1 implies that $\eta^{**}_{III} > 0$. To see this, note that given the condition,

η^{**}_{III} can be written as:

$$\eta^{**}_{III} \geq \frac{\theta}{\delta}\left(\overline{\gamma} - \beta^{**}_{III}\right) - \left(\frac{1}{2}\overline{\gamma} - \beta^{**}_{III}\right) b^{**}_{III} > 0, \tag{3.40}$$

where the sign follows from $\theta \geq 1$, $0 < \delta < 1$, and $0 < b^{**}_{III} < 1$.

Finally, recall that case III applies under the assumption that $\beta > 0$. Rewriting $\beta^{**}_{III} > 0$ yields that this assumption only holds if:

$$b^{**}_{III} = 1 - \theta\frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta} > \frac{\delta}{1-\delta}\overline{U_A}. \tag{3.41}$$

In turn, condition (3.41) is satisfied if and only if $\delta \in \left(\delta^L, \delta^M\right)$, where the threshold values of the discount factor are defined as:

$$\delta^L \equiv \frac{1 + \theta\overline{\gamma} - \sqrt{1 - 2\overline{U_A}\theta\overline{\gamma}}}{2 + \theta\overline{\gamma} + 2\overline{U_A}}, \text{ and} \tag{3.42}$$

$$\delta^M \equiv \frac{1 + \theta\overline{\gamma} + \sqrt{1 - 2\overline{U_A}\theta\overline{\gamma}}}{2 + \theta\overline{\gamma} + 2\overline{U_A}}. \tag{3.43}$$

Note that $\delta^L > 0$ and $\delta^L < \delta^M$ because $1 + \theta\overline{\gamma} > \sqrt{1 - 2\overline{U_A}\theta\overline{\gamma}} > 0$. The first inequality follows from $\overline{U_A} > 0$, $\theta > 1$, and $\overline{\gamma} > 0$. The second inequality follows from assumption (3.4) and $0 < \overline{\gamma} < 1$.

Cases I and II. Let's now assume that the condition on β binds, that is, $\beta = 0$ and $\lambda_4 > 0$. If $\beta = 0$, there are two subcases. The first subcase is where the altruistic supervisor's truthtelling constraint (3.24) binds, that is, $\lambda_2 > 0$. This leads to what has been labeled case II in the main text. In solving this subcase, it is assumed that also the selfish supervisor's truthtelling constraint (3.25) binds, that is, $\lambda_3 > 0$. Along the way, it is shown that this assumption indeed holds.

Solving the supervisor types' truthtelling constraints (3.24 and 3.25), given $\beta = 0$, yields:

$$b^{**}_{II} = \frac{\delta}{1-\delta}U_A, \text{ and} \tag{3.44}$$

$$\eta^{**}_{II} = \frac{\theta\left(1 - 2\delta + \delta^2\right)\left(\overline{U_S} - \alpha_{II}\right) + \overline{\gamma}\delta^2 U_A^2}{\delta\left(1-\delta\right)U_A}. \tag{3.45}$$

One can now compute the supervisor's compensation to be:

$$w_{S,II}^{**} = \overline{U_S}. \tag{3.46}$$

Given the solutions found so far, the first-order conditions (3.32), (3.33), and (3.35) can be solved for the Lagrange multipliers. This yields:

$$\lambda_1 = \frac{1}{\delta(1-\delta)\theta}\left[\frac{\delta}{1-\delta}U_A - 1 + \theta\frac{1-\delta}{\delta}\right] > 0, \tag{3.47}$$

$$\lambda_2 = \frac{1}{(1-\delta)\overline{\gamma}\theta}\left[1-\delta(1+U_A)\right], \text{ and} \tag{3.48}$$

$$\lambda_3 = \frac{1}{\overline{\gamma}\theta}\left[\frac{\delta}{1-\delta}U_A - 1 + \theta\frac{1-\delta}{\delta}\overline{\gamma}\right] > 0. \tag{3.49}$$

The signs of λ_1 and λ_3 follow from the assumptions that $\beta = 0$, which has been shown above to imply that $\frac{\delta}{1-\delta}\overline{U_A} - 1 + \theta\frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta} > 0$ (see condition (3.41)), and $\overline{\gamma} < 1$. Moreover, since $\lambda_1 > 0$, the agent's participation constraint (3.23) binds, from which it follows that $a_{II}^{**} = \overline{U_A} - \frac{1}{2}\frac{\left(b_{II}^{**}\right)^2}{\theta}$ and $U_A = \overline{U_A}$. The proof that $\eta_{II}^{**} > 0$ is analogous to the one given in case III. Proceeding in the same way as there, one can check that:

$$\eta_{II}^{**} \geq \frac{\overline{\gamma}}{2\delta(1-\delta)}\left[2\theta(1-\delta) + \delta^2\overline{U_A}\right] > 0. \tag{3.50}$$

To conclude case II, recall that it has been assumed that $\lambda_2 > 0$ and $\lambda_3 > 0$. The latter assumption indeed holds, as shown in (3.49). However, inspecting (3.48) reveals that the former assumption only holds if $\delta < \delta^H$, where the threshold value of the discount factor is defined as:

$$\delta^H \equiv \frac{1}{1+\overline{U_A}}. \tag{3.51}$$

One can check that $\delta^M < \delta^H$. This inequality follows from $\theta > 1$, $\overline{U_A} > 0$, and $\overline{\gamma} > 0$. Therefore, case II applies for values of the discount factor such that $\delta \in (0, \delta^L]$ and $\delta \in [\delta^M, \delta^H)$.

The second subcase is where the altruistic supervisor's truthtelling constraint (3.24) is slack, that is, $\lambda_2 = 0$. As shown above, this holds for values of the discount

factor $\delta > \delta^H$. The set of contracts that results has been labeled case I in the main text. Solving the first-order conditions (3.32), (3.33), and (3.35) yields $\lambda_1 = 1 > 0$, $\lambda_3 = \frac{1-\delta}{\delta} > 0$, and $b_I^{**} = 1$. From $\lambda_1 > 0$ it follows that the agent's participation constraint (3.23) binds, implying that $a_I^{**} = \overline{U_A} - \frac{1}{2\theta}$ and $U_A = \overline{U_A}$. From $\lambda_3 > 0$ it follows that the selfish supervisor type's truthtelling constraint (3.25) binds. Rewriting (3.25), given that $\beta_I^{**} = 0$, yields:

$$\eta_I^{**} = \theta \left(\overline{U_S} - \alpha_I \right). \tag{3.52}$$

One can now compute the supervisors's compensation to be:

$$w_{S,I}^{**} = \overline{U_S}. \tag{3.53}$$

The proof that $\eta_I^{**} > 0$ is analogous to the one given in case III. Proceeding in the same way as there, one can check that:

$$\eta_I^{**} \geq \frac{1}{2}\overline{\gamma} + \theta\overline{\gamma}\left(1 + \overline{U_A}\right) > 0. \tag{3.54}$$

By conclusion, case III applies for $\delta \in \left(\delta^L, \delta^M\right)$, case II applies for $\delta \in (0, \delta^L]$ and $\delta \in [\delta^M, \delta^H)$, and case I applies for the remaining values of the discount factor, $\delta \in [\delta^H, 1)$. Also, it has been shown to hold that $0 < \delta^L < \delta^M < \delta^H < 1$.

3.B Allowing for objective incentives for the agent

In this Appendix I show that, as long as σ_μ^2 is sufficiently large, allowing for objective incentives for the agent does not change the key results derived in this chapter. The way I model the verifiable information about the agent's performance is identical to Baker et al. (1994).

Suppose that the agent's contract also contains a bonus v to be paid if the verifiable signal is high ($z = 1$). Given that the supervisor reports performance truthfully, and some realization of μ which the agent has privately observed, the agent's optimal effort choice is:

$$e = \frac{b + \mu v}{\theta}, \tag{3.55}$$

which varies with μ. The agent's participation constraint now reads:

$$E\left[a + \frac{b + \mu v}{\theta}(b + \mu v) - \frac{1}{2}\theta\left(\frac{b + \mu v}{\theta}\right)^2\right] \geq \overline{U_A}, \tag{3.56}$$

where the expectation runs over μ. Using $E[\mu] = 1$ and $E[\mu^2] = 1 + \sigma_\mu^2$, the participation constraint can be written as:

$$a + \frac{1}{2}\frac{b^2 + 2bv + \left(1 + \sigma_\mu^2\right)v^2}{\theta} \geq \overline{U_A}. \tag{3.57}$$

Rewriting (3.57) in another way yields that the agent must be paid at least a total expected compensation level of:

$$w_A \geq \overline{U_A} + \frac{1}{2}\frac{b^2 + 2bv + \left(1 + \sigma_\mu^2\right)v^2}{\theta}. \tag{3.58}$$

Equation (3.58) reveals that, if $\sigma_\mu^2 > 0$, the agent's compensation rises faster in v compared to b. The reason is as follows. If $v > 0$, the agent will sometimes exert a high effort level and sometimes a low effort level depending on the realization of μ. Because effort costs are convex, average effort costs increase in the variation in e, and therefore in σ_μ^2. For these costs the agent must be compensated. Using objective incentives for the agent thus generates its own kind of agency costs. I assume that using only objective incentives for the agent does not generate sufficient value to

compensate the agent for his reservation utility. One can show that this assumption holds if:

$$\sigma_\mu^2 > \frac{1 - 2\theta\overline{U_A}}{2\theta\overline{U_A}} > 0, \tag{3.59}$$

where the sign follows from assumption (3.4). In the following I derive the optimal contracts in case the principal can also use objective incentives for the agent.

First, note that the sets of contracts derived in case the principal screens (Proposition 3.1) and in case the principal attracts both supervisor types and $\delta \geq \delta^H \equiv \frac{1}{1+\overline{U_A}}$ (Proposition 3.2, case I), are still optimal. The reason is that given these sets of contracts, the supervisor nor the agent earns a rent, and the agent is induced to exert the first-best effort level. It follows that there is no way to improve upon these sets of contracts by using objective incentives for the agent. This observation already enables us to to confirm some key results. First, given the contracts from Proposition 3.2, case I, the supervisor's altruism provides a net incentive to report performance truthfully, rather than to bias performance evaluations (Corollary 1). Second, the proof that screening is optimal for high values of the discount factor, presented in subsection 3.4.4, depends only on the aforementioned sets of contracts. Since these sets of contracts do not change, it follows that this result is insensitive to allowing for objective incentives for the agent. In the remainder of this Appendix, I derive the optimal contracts in case the principal attracts both supervisor types and it holds that $\delta < \delta^H$. The derivations are largely similar to the ones in the proof of Proposition 3.2.

Given the agent's effort e (3.55), $E[\mu] = 1$, and $E[\mu^2] = 1 + \sigma_\mu^2$, the altruistic and selfish supervisor will report performance truthfully if:

$$\frac{\delta}{1-\delta}\left[\alpha + \frac{(b+v)(\eta - b\beta) + v\sigma_\mu^2\eta}{\theta} + \overline{\gamma}U_A - \overline{U_S}\right] \geq |\overline{\gamma} - \beta|\, b, \text{ and} \tag{3.60}$$

$$\frac{\delta}{1-\delta}\left[\alpha + \frac{(b+v)(\eta - b\beta) + v\sigma_\mu^2\eta}{\theta} - \overline{U_S}\right] \geq |0 - \beta|\, b, \tag{3.61}$$

respectively. Further, recall from the proof of Proposition 3.2 that it must hold that:

$$\beta \geq 0. \tag{3.62}$$

The principal's problem can now be written as:

$$\begin{aligned}\max_{a,b,v,\beta,\eta} \quad \Gamma \;=\; & \frac{b+v}{\theta} - \left[a + \frac{b^2 + 2bv + \left(1+\sigma_\mu^2\right)v^2}{\theta}\right] - \\ & \left[\alpha + \frac{(b+v)(\eta - b\beta) + v\sigma_\mu^2\eta}{\theta}\right], \\ & \text{s.t. (3.57), (3.60), (3.61), and (3.62).}\end{aligned}$$

The first-order conditions to the problem are:

$$\frac{\partial \Gamma}{\partial a} = -1 + \lambda_1 + \lambda_2 \frac{\delta}{1-\delta}\overline{\gamma} = 0, \tag{3.63}$$

$$\begin{aligned}\frac{\partial \Gamma}{\partial b} = & \frac{1}{\theta} - \frac{2b+2v}{\theta} - \left[\frac{1}{\theta}(\eta - \beta b) - \beta\frac{b+v}{\theta}\right] + \lambda_1\frac{b+v}{\theta} + \\ & \lambda_2\left\{\frac{\delta}{1-\delta}\left[\frac{\eta-\beta b}{\theta} - \beta\frac{b+v}{\theta} + \overline{\gamma}\frac{b+v}{\theta}\right] - |\overline{\gamma} - \beta|\right\} + \\ & \lambda_3\left\{\frac{\delta}{1-\delta}\left[\frac{\eta-\beta b}{\theta} - \beta\frac{b+v}{\theta}\right] - |0-\beta|\right\} = 0,\end{aligned} \tag{3.64}$$

$$\begin{aligned}\frac{\partial \Gamma}{\partial v} = & \frac{1}{\theta} - \frac{2b + 2\left(1+\sigma_\mu^2\right)v}{\theta} - \frac{\eta\left(1+\sigma_\mu^2\right) - b\beta}{\theta} + \lambda_1\frac{b + v\left(1+\sigma_\mu^2\right)}{\theta} + \\ & \lambda_2\frac{\delta}{1-\delta}\left[\frac{\eta\left(1+\sigma_\mu^2\right) - b\beta}{\theta} + \overline{\gamma}\frac{b + v\left(1+\sigma_\mu^2\right)}{\theta}\right] + \lambda_3\frac{\delta}{1-\delta}\frac{\eta\left(1+\sigma_\mu^2\right) - b\beta}{\theta} = 0,\end{aligned} \tag{3.65}$$

$$\begin{aligned}\frac{\partial \Gamma}{\partial \beta} = & \frac{b+v}{\theta}b - \lambda_2\left(\frac{\delta}{1-\delta}\frac{b+v}{\theta}b + \frac{\beta - \overline{\gamma}}{|\overline{\gamma} - \beta|}b\right) - \\ & \lambda_3\left(\frac{\delta}{1-\delta}\frac{b+v}{\theta}b + \frac{\beta - 0}{|0-\beta|}b\right) + \lambda_4 = 0, \text{ and}\end{aligned} \tag{3.66}$$

$$\frac{\partial \Gamma}{\partial \eta} = -\frac{b + \left(1+\sigma_\mu^2\right)v}{\theta} + \lambda_2\frac{\delta}{1-\delta}\frac{b + \left(1+\sigma_\mu^2\right)v}{\theta} + \lambda_3\frac{\delta}{1-\delta}\frac{b + \left(1+\sigma_\mu^2\right)v}{\theta} = 0, \tag{3.67}$$

where λ_1, λ_2, λ_3, and λ_4 are the Lagrange multipliers on (3.57), (3.60), (3.61), and (3.62), respectively. As in Proposition 3.2, if $\delta < \delta^H$, two cases can be distinguished. Since the cases are closely comparable to cases II and III from Proposition 3.2, they are again denoted by subscripts II and III. I start by solving case III.

Case III. Let's first assume that condition (3.62) does not bind, that is, $\beta > 0$ and $\lambda_4 = 0$. Solving (3.63), (3.66), and (3.67) then yields $\lambda_1 = 1 - \frac{1}{2}\overline{\gamma} > 0$ and

$\lambda_2 = \lambda_3 = \frac{1-\delta}{2\delta} > 0$. Hence, the agent's participation constraint (3.57) binds. This implies that $U_A = \overline{U_A}$ and $a^{**}_{III} = \overline{U_A} - \frac{1}{2}\frac{b^2_{III}+2b_{III}v_{III}+\left(1+\sigma^2_\mu\right)v^2_{III}}{\theta}$. Further, both supervisor types' truthtelling constraints (3.60 and 3.61) also bind. Solving these constraints yields:

$$\beta^{**}_{III} = \frac{\overline{\gamma}}{2}\left(1 - \frac{\delta}{1-\delta}\frac{\overline{U_A}}{b_{III}}\right), \text{ and} \tag{3.68}$$

$$\eta^{**}_{III} = \frac{\theta\left(\frac{1-\delta}{\delta}\beta^{**}_{III}b_{III} + \overline{U_S} - \alpha^{**}_{III}\right) + \left(b_{III}+v_{III}\right)b_{III}\beta^{**}_{III}}{b_{III} + \left(1+\sigma^2_\mu\right)v_{III}}, \tag{3.69}$$

where α^{**}_{III} must satisfy Lemma 3.1. Next, solving (3.64) and (3.65) for b_{III} and v_{III} yields:

$$b^{**}_{III} = 1 - \theta\frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta}\frac{1+\sigma^2_\mu}{\sigma^2_\mu}, \text{ and} \tag{3.70}$$

$$v^{**}_{III} = \frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta}\frac{\theta}{\sigma^2_\mu} > 0. \tag{3.71}$$

Thus, in case III the principal optimally uses objective incentives for the agent in order to reduce the rents the supervisor requires to report performance truthfully.

To conclude case III, recall that it applies as long as it holds that $\beta^{**}_{III} > 0$. This inequality can be written as:

$$b^{**}_{III} = 1 - \theta\frac{\overline{\gamma}}{2}\frac{1-\delta}{\delta}\frac{1+\sigma^2_\mu}{\sigma^2_\mu} > \frac{\delta}{1-\delta}\overline{U_A}. \tag{3.72}$$

Solving the above condition yields that it must hold that $\delta \in \left(\delta^L, \delta^M\right)$, where the threshold values of the discount factor are given by:

$$\delta^L \equiv \frac{\sigma^2_\mu + \overline{\gamma}\theta\left(1+\sigma^2_\mu\right) - \sqrt{\sigma^2_\mu\left[\sigma^2_\mu - 2\overline{U_A}\overline{\gamma}\theta\left(1+\sigma^2_\mu\right)\right]}}{2\sigma^2_\mu + \overline{\gamma}\theta\left(1+\sigma^2_\mu\right) + 2\overline{U_A}\sigma^2_\mu}, \text{ and} \tag{3.73}$$

$$\delta^M \equiv \frac{\sigma^2_\mu + \overline{\gamma}\theta\left(1+\sigma^2_\mu\right) + \sqrt{\sigma^2_\mu\left[\sigma^2_\mu - 2\overline{U_A}\overline{\gamma}\theta\left(1+\sigma^2_\mu\right)\right]}}{2\sigma^2_\mu + \overline{\gamma}\theta\left(1+\sigma^2_\mu\right) + 2\overline{U_A}\sigma^2_\mu}. \tag{3.74}$$

Note that δ^L and δ^M only exist if the root terms in the expressions above are well-defined. This is only the case if $\sigma^2_\mu - 2\overline{U_A}\overline{\gamma}\theta\left(1+\sigma^2_\mu\right) > 0$, which can be written

as:

$$\sigma_\mu^2 > \frac{2\overline{U_A}\overline{\gamma}\theta}{1-2\overline{U_A}\overline{\gamma}\theta}. \tag{3.75}$$

The above inequality is not necessarily implied by the assumption on the minimum level of σ_μ^2 (assumption (3.59)). It follows that case III may exist or not exist, depending on the level of σ_μ^2.

Case II. The next step is to derive the optimal contracts for $\delta \notin \left(\delta^L, \delta^M\right)$, given that $\delta < \delta^H$ (case II). This case always exists. To see this, first note that if δ^L and δ^M are not well-defined, case II applies for all $\delta < \delta^H$. Second, given that it is well-defined, it can be easily checked that $\delta^L > 0$. For the remainder of the proof, it does not matter whether or not $\delta^M < \delta^H$.

First note that a property of case II is that $\beta_{II}^{**} = 0$, since for any $\beta > 0$ case III is optimal. Next, case II can be easily solved by the following argument. Let's suppose that the agent's participation constraint (3.57) and both supervisor types' truthtelling constraints (3.60 and 3.61) bind. If the agent's participation constraint (3.57) binds, it holds that $U_A = \overline{U_A}$ and $a_{II}^{**} = \overline{U_A} - \frac{1}{2}\frac{b_{II}^2 + 2b_{II}v_{II} + \left(1+\sigma_\mu^2\right)v_{II}^2}{\theta}$. Moreover, given $\beta_{II}^{**} = 0$, solving both supervisor types' truthtelling constraints (3.60 and 3.61) yields:

$$b_{II}^{**} = \frac{\delta}{1-\delta}\overline{U_A}, \text{ and} \tag{3.76}$$

$$\eta_{II}^{**} = \frac{\theta\left(\overline{U_S} - \alpha_{II}^{**}\right)}{b_{II}^{**} + \left(1+\sigma_\mu^2\right)v_{II}}, \tag{3.77}$$

where α_{II}^{**} must satisfy Lemma 3.1. Note that $b_{II}^{**} < 1$ by the assumption that $\delta < \delta^H$. Next, for any base salary α_{II}^{**} that satisfies Lemma 3.1 and the solutions found so far, the supervisor's compensation can be computed to be $w_{S,II}^{**} = \overline{U_S}$. The supervisor cannot be paid a lower compensation level, since otherwise the selfish type will reject the job. It follows directly that it is indeed optimal to have the agent's participation constraint (3.57) and both supervisor types' truthtelling constraints (3.60 and 3.61) bind. The reason is that the agent earns the lowest possible compensation level (since (3.57) binds), the supervisor earns the lowest possible compensation level, and the agent receives the highest possible subjective bonus (note that b_{II}^{**} cannot

be greater than $\frac{\delta}{1-\delta}\overline{U_A}$, as this would imply that case III is optimal, see condition (3.72)).

Importantly, in case II the optimal contracts imply that supervisor's altruism provides a net incentive to report performance truthfully (Corollary 1). To see this, note that given $\beta_{II}^{**} = 0$ and $w_{S,II}^{**} = \overline{U_S}$, the altruistic type's truthtelling constraint (3.60) reduces to $\frac{\delta}{1-\delta}\overline{U_A} \geq b_{II}^{**}$, which always holds.

Finally, one can now derive the optimal value of the objective bonus v_{II}. Using $E[\mu] = 1$, $E[\mu^2] = 1 + \sigma_\mu^2$, and the solutions found so far, the optimal value of v is given by:

$$\max_{v_{II}} \quad E[\Pi_{II}^{**}] = \frac{\frac{\delta}{1-\delta}\overline{U_A} + v_{II}}{\theta} - \overline{U_S} - \left[\overline{U_A} + \frac{1}{2}\frac{\left(\frac{\delta}{1-\delta}\overline{U_A}\right)^2 + 2\frac{\delta}{1-\delta}\overline{U_A}v_{II} + \left(1+\sigma_\mu^2\right)v_{II}^2}{\theta}\right], \tag{3.78}$$

which yields:

$$v_{II}^{**} = \frac{1 - \frac{\delta}{1-\delta}\overline{U_A}}{1+\sigma_\mu^2} > 0, \tag{3.79}$$

where the sign follows from $b_{II}^{**} = \frac{\delta}{1-\delta}\overline{U_A} < 1$. Hence, in case II, the principal avoids paying the supervisor rents, but at the costs of setting the subjective bonus below the first-best level. To get the agent's effort level closer to the first-best, the principal adds objective incentives to the agent's contract.

It is now possible to confirm the final key result, namely that screening is optimal for relatively low values of the discount factor. The proof is analogous to the one given in subsection 3.4.4. Recall that the principal's per-period utility from screening, conditional on the supervisor and agent accepting their contract, equals $\frac{1}{2\theta} - (1-\overline{\gamma})\overline{U_A} - \overline{U_S}$. This profit is strictly positive by assumption (3.4) and $\overline{\gamma} > 0$. Next, suppose that $\delta \to 0$. For $\delta = 0$ it holds that the subjective bonus equals $b_{II}^{**} = 0$. By assumption (3.59), the principal's utility is strictly negative if she only uses an objective incentive scheme. It follows that there must be a range of discount factors $\delta \in (0, \underline{\delta})$ for which screening is optimal, where $\underline{\delta} > 0$.

Chapter 4

Endogenous Effort Norms in Hierarchical Firms

4.1 Introduction

Empirical research suggests that employees have a preference to conform to social norms within their firm, even if this requires taking costly actions.[1] Given that employees conform to social norms, firms have an incentive to affect these norms. After all, a firm only benefits from employee conformism to the extent that the existing norms prescribe to act in the firm's interest. From a practical point of view, an important question is therefore how firms can increase profits by creating efficient norms. From a theoretical point of view, an important question is whether and how firms' desire to create efficient norms can explain firm behavior and organization. This chapter develops and analyzes a theoretical model that yields answers to both questions.

The novelty of the present analysis is its focus on a particular determinant of norms for employees, namely the actual behavior of superiors. Apart from being intuitive, there is empirical evidence suggesting that the example of superiors has normative implications for subordinates. Treviño et al. (1999) find a strong nega-

[1] Important studies documenting this phenomenon are Ichino and Maggi (2000), Falk and Ichino (2006), Bradley et al. (2007), Mas and Moretti (2009), and Bradler et al. (2013). See the next section for a detailed discussion of these papers.

tive correlation between employees' perceptions of, on the one hand, the quality of ethical leadership of executives and supervisors within their firm, and on the other hand, unethical behavior within their firm (also see Treviño and Weaver 2003, chapter 9, where similar evidence is reported). In line with this, Posner and Schmidt (1992) document that 92% of American managers agree with the statement that "the behavior of those in charge is the *principle* determinant of the 'ethical tone' of my firm". This view is confirmed by J. Irwin Miller, a successful CEO himself, who claims that "all of the corporate standards of ethics don't mean anything unless the persons in the corporation perceive the top people to abide by them when the going is really tough" (quoted from Murphy and Enderle 1995).

The formal model studied below considers a hierarchical firm that consists of three layers: a residual claimant principal, a supervisor, and an agent. All players exert non-verifiable effort which yields valuable production for the principal's firm. Both employees (the supervisor and the agent) incur a psychological cost if their effort falls short of their norm for effort, as in e.g. Fischer and Huddart (2008). To reflect the evidence cited above, employees consider the example of their superior to be the norm for their own effort. Hence, the principal's effort constitutes the norm for the supervisor, and the supervisor's effort constitutes the norm for the agent. The remaining part of the Introduction describes the results of the chapter.

The key result from the analysis is that, in equilibrium, norms are lower as one moves down the hierarchy. I call this phenomenon 'norm erosion'. Norm erosion arises because when exerting effort, employees take into account the example of their superior, but never fully comply with it. The reason is simple: employees find it privately optimal to trade off some costs of performing below the norm against the benefit of avoiding some effort costs. It follows that each employee sets an example for his subordinate that is lower than the example he got from his superior. The upshot from this result is that effort norms within firms are not only determined by the example set at the top of the hierarchy, but also by how this example is transmitted to lower organizational layers. Since firms do not control this process, they incur a number of costs when creating optimal effort norms for employees, as explained next.

To understand why norm erosion is costly to firms, one needs to know that there is a unique first-best effort norm for employees. This holds because a higher norm induces more valuable effort from the employee, but also imposes more effort costs and norm violation costs upon the employee for which he must be compensated through his salary. The first-best norm optimally trades off these marginal costs and benefits. Next, when choosing how much effort to exert, the principal has two objectives in mind. On the one hand, she wants to set the first-best norm for her employees. On the other hand, she wants to maximize her own contribution to firm profits. The effort level that maximizes the principal's own contribution to firm profits is referred to as the principal's first-best effort level. Under a mild assumption that I use, it holds that the principal's first-best effort level is identical to the first-best norm for both employees. It follows that if there were no norm erosion, the principal would maximize total profits by exerting her first-best effort level. However, knowing that the supervisor will erode her example, the principal optimally sets a higher example to begin with. In this way the principal makes sure that the agent's equilibrium norm is closer to its first-best level. The cost of this strategy is two-fold: the principal works inefficiently hard, and the supervisor faces an inefficiently high norm. Compared to the first-best case, hierarchical firms are thus confronted with three inefficiencies in norm creation: the principal's effort is too high, the supervisor's norm is too high, and the agent's norm is too low. The remaining results show that, to minimize these inefficiencies, hierarchical firms optimally adjust the way they are organized.

First, hierarchical firms optimally promote employees with the strongest sensitivity to social norms. A supervisor who is more sensitive to norms will conform to the principal's example to a larger extent. As a result, norms are eroded to a lower extent, which is always valuable to the principal. The reason is that establishing a given effort norm for the agent now requires a lower effort level from the principal. This is profitable since in equilibrium the principal works inefficiently hard and the supervisor's norm is inefficiently high. Alternatively, for a given effort level of the principal and norm for the supervisor, the agent will be faced with a higher norm. This is also profitable because in equilibrium the agent's norm is inefficiently low.

Second, hierarchical firms optimally distort the supervisor's span of control, that is, the number of agents heading under one supervisor. The analysis shows that firms may set the supervisor's span of control both above and below its first-best level. The mechanism that produces these results is the following. In the model I assume that for supervision technology reasons, there is some first-best supervisor span of control. When choosing how many supervisors and agents to hire, the principal thus has an incentive to stick as closely as possible to this exogenously given span of control. However, the firm's profits also depend on which kind of employee faces the most efficient effort norm in equilibrium. Recall that the equilibrium norm for supervisors is always above the first-best level, whereas the equilibrium norm for agents is always below the first-best level. Yet, the extent to which equilibrium norms diverge from the first-best level may be different for agents and supervisors. In fact, the principal optimally makes sure that the kind of employee that is relatively abundant faces the more efficient norm. The principal does this by raising her own effort level in the number of agents, and by decreasing her effort in the number of supervisors. Therefore, given that the supervision technology is such that it is attractive to hire relatively many agents (supervisors), the principal optimally makes sure that the effort norm for agents (supervisors) is more efficient than the effort norm for supervisors (agents), which in turn makes it attractive to hire even more agents (supervisors). As a result, the principal optimally distorts the supervisor's span of control away from its first-best level.

Last, hierarchical firms suffer from norm erosion precisely because they are hierarchical. A straightforward prediction following from the model is therefore that firms optimally keep the extent of hierarchy to a minimum.

This chapter proceeds as follows. The next section describes the related literature. Section 4.3 develops the formal model. Section 4.4 solves the model, while section 4.5 explores implications for organizational design. Section 4.6 finishes with concluding remarks.

4.2 Related literature

Empirical research suggests that employees have a preference to conform to social norms within their firm. Ichino and Maggi (2000) show that employees engage less in misconduct and absent themselves less often if their co-workers are less likely to do so (also see Bradley et al. 2007). Falk and Ichino (2006) report that, if employees work in pairs rather than individually, the standard deviation of output is smaller and output is higher. Moreover, the authors find that low-productivity employees respond strongest to working in pairs. Mas and Moretti (2009) find that supermarket cashiers improve their performance if a high-productivity cashier enters their shift. As in Falk and Ichino (2006), this effect is far more pronounced for low-productivity workers than for high-productivity workers. Mas and Moretti (2009) provide a social pressure interpretation for their results, based on the finding that cashiers only improve their performance if this can be observed by the new cashier. Finally, in a controlled work environment, Bradler et al. (2013) show that the provision of public recognition to employees improves their performance. Similar to Falk and Ichino (2006) and Mas and Moretti (2009), the authors find that their results are mainly driven by those employees who did not receive recognition, and thus learned that they performed worse than others. For this reason, Bradler et al. (2013) explain their results partly in terms of employees' preference to conform to a group norm.

An early theoretical paper studying social norms is Akerlof (1980). Akerlof's (1980) model shows that an existing norm may stay in place if deviating from the norm leads to a loss in reputation (also see Bernheim 1994, who assumes that people care for others' perceptions of one's preferences). Another early contribution is the paper by Kandel and Lazear (1992), who study peer pressure as a mechanism that may reduce free-riding problems in team production.

More recently, Sliwka (2007) develops a model where agents may prefer to conform their effort to the social norm, but are uncertain about what the social norm prescribes. This opens the door for the principal to affect the agent's perception of the social norm through her own actions. Specifically, Sliwka (2007) shows that offering incentives may be a credible signal that the social norm is to act selfishly,

whereas offering no incentives may be a credible signal that the social norm is to act fairly, that is, to exert at least some effort. Fischer and Huddart (2008) study firms where norms exist for desirable actions (like exerting effort) and undesirable actions (like earnings management). The authors show that social norms multiply the impact of individual incentives on agents' behavior. The reason is that individual incentives change an individual agent's behavior, but this also affects the social norm to which all agents want to conform. Fischer and Huddart (2008) also show that it may be optimal to split firms in parts. The benefit is that having separate departments allows for the cultivation of different social norms for different tasks.

Stevens and Thevaranjan (2010) study optimal efficiency wage contracts for morally sensitive agents. The authors assume that the principal can declare what the effort norm is at the contracting stage. Stevens and Thevaranjan (2010) show that, when hiring a morally sensitive agent, the principal may achieve first-best profits even when she does not use incentives. Huck et al. (2012) define social norms in terms of Pareto efficiency. Given this definition, the authors demonstrate that a social norm makes team-incentives more effective. The reason is that team-incentives create positive externalities among employees, which leads to a higher social norm for effort. The opposite reasoning applies to using relative incentives like tournaments. Huck et al. (2012) further show that, if team incentives are used, social norms may give rise to multiple equilibria, some of which may lead to inefficiently high social norms.

Akerlof and Kranton (2005) introduce the notion that a person's utility may depend on his or her social identity. In their model, a social identity exogenously implies some norm for behavior. Akerlof and Kranton (2005) show that employees identifying with the firm accept lower wages and require lower incentives. The authors consequently claim that it is valuable for firms to invest in changing their employees' identity (also see Heinle et al. 2012, who derive optimal contracts when agents may identify with the firm in a multi-task setting). Carlin and Gervais (2009) model morality as a self-imposed restriction to exert high effort. In their analysis, norms are thus a purely personal trait that cannot be affected by the firm in any way. The authors derive a number of predictions as to how the presence of virtuous

agents in the labor market affect firms' optimal contract design, project choice, and extent of bureaucracy.

In contrast to all papers above, the starting point of this chapter is that the actual example of superiors determines effort norms for employees. Norms are thus not exogenously given (as in Akerlof and Kranton 2005, Carlin and Gervais 2009, and Heinle et al. 2012), but depend on the behavior of superiors. Also, norms cannot be costlessly declared (as in Stevens and Thevaranjan 2010), but creating norms requires setting a costly example. Finally, the principal does not only indirectly influence norms through making institutional choices (as in Sliwka 2007, Fischer and Huddart 2008, and Huck et al. 2012), but also directly through setting her own example.

More distantly related to the present paper is Hermalin's (1998) analysis of leadership. Hermalin (1998) studies how a leader can credibly communicate information to his team-members about the marginal productivity of their effort. One of the mechanisms Hermalin (1998) considers is 'leading by example'. That is, if the leader exerts a high effort level and thereby incurs high effort costs, this credibly signals to the other team-members that exerting effort is valuable. The parallel between Hermalin (1998) and the present analysis is the importance of the leader's example for the behavior of others. However, different problems are studied. In Hermalin's (1998) analysis the credible transmission of information is at stake, whereas here the principal sets her example so as to create optimal effort norms for her employees.

Finally, another important difference with the existing literature is that the present analysis considers hierarchical firms that consist of at least three layers. Hence, I do not only study how the example at the top of the hierarchy matters for effort norms, but also the process of transmitting this example to lower layers of the hierarchy. In fact, the key result of the chapter is to show that this process is inefficient from the firm's perspective.

4.3 A model of effort norms in hierarchical firms

Consider a hierarchical firm owned by one principal. The principal's organizational layer is denoted by $l = 0$. In the most basic specification of the model, the principal hires one supervisor ($l = 1$) and one agent ($l = 2$) (subsection 4.5.2 extends the model to the case where the principal hires S supervisors and A agents). All players exert unverifiable effort e_l, which yields valuable production for the principal's firm. The marginal product of effort is given by p_l. I assume that $p_l \geq p_{l+1}$, implying that a player in a higher layer of the organization has a weakly higher marginal productivity. The costs of exerting effort are given by $\frac{1}{2}\theta e_l^2$.

Employee utility from the job is given by:

$$U_l = w_l - \frac{1}{2}\theta e_l^2 - \frac{1}{2}\gamma \left(n_l - e_l\right)^2, \tag{4.1}$$

where w_l denotes a fixed wage (in Appendix 4.A, I show that as long as employees are risk averse, the results of the chapter are insensitive to the use of performance pay). The last term in (4.1) describes the employee's preference for conforming to social norms, which consists of two parts. The term $n_l - e_l$ describes how much the employee's effort differs from the norm for effort, n_l. The parameter γ reflects the strength of the employee's sensitivity to norms. In the basic model, γ is assumed to be identical for all employees (section 4.5.1 extends the model to the case where employees differ in norm sensitivity). Throughout the chapter, I assume that the principal can observe γ. The outside option utility is assumed to be identical for all employees, and equal to $u = 0$.

The novelty of the model is how social norms for effort are determined. Formally, the effort norm for an employee in organizational layer l is given by:

$$n_l = e_{l-1}\frac{p_l}{p_{l-1}}. \tag{4.2}$$

The effort norm consists of two parts. First, as laid out in the Introduction, the actual behavior of superiors plays an important role in shaping norms within organizations. Hence, I assume that the supervisor's norm is affected by the principal's

effort, and the agent's norm is affected by the supervisor's effort.[2] Note that it is implicitly assumed that employees can observe their superior's effort level. Second, I assume that employees are aware of the fact that their superior's effort may be more valuable than their own ($p_{l-1} > p_l$). If this is the case, from an efficiency point of view, the superior should work harder than the employee. For this reason, I assume that employees do not consider their superior's effort to be normative to the extent that the superior's marginal productivity is higher. This behavioral assumption implies that, qualitatively, the results of the chapter do not depend on differences in marginal productivities p_l. For convenience, I therefore solve the model assuming that marginal productivities are identical across organizational layers, that is, $p_l = p$ for all l. It follows that norms are given by:

$$n_l = e_{l-1}. \tag{4.3}$$

The order of the game is as follows. First, the principal exerts an effort level, and thereby sets an example. Second, the principal hires at least one supervisor and one agent, and offers all employees a compensation contract. Third, the employees decide whether or not to accept the contract. If one of them rejects the contract, the game ends. Last, if all employees accepted the contract, they exert effort and payoffs realize. In the next section, I solve for the subgame perfect equilibrium of the game by using backwards induction.

4.4 Analysis

4.4.1 Contractible effort

In this subsection, I derive the first-best benchmark where the employees' effort is verifiable and thus contractible. First note that, to satisfy the employees' participa-

[2] In general, the agent's norm may also directly be affected by the example of the principal. However, this would not affect the results of the chapter qualitatively, as long as the agent attaches a non-negative weight to the example of the supervisor. This is not a strong assumption, since supervisors typically represent the firm and its policies toward their subordinates.

tion constraint, they must be paid a minimum salary equal to:

$$w_l = \frac{1}{2}\theta e_l^2 + \frac{1}{2}\gamma\left(n_l - e_l\right)^2. \tag{4.4}$$

Given the lowest possible salary (4.4), and taking into account that the norm for layer l is given by the effort level exerted in layer $l-1$, firm profits can be written as:

$$\Pi = \underbrace{pe_0 - \frac{1}{2}\theta e_0^2}_{\text{principal } (l=0)} + \underbrace{pe_1 - \frac{1}{2}\theta e_1^2 - \frac{1}{2}\gamma\left(e_0 - e_1\right)^2}_{\text{supervisor } (l=1)} + \underbrace{pe_2 - \frac{1}{2}\theta e_2^2 - \frac{1}{2}\gamma\left(e_1 - e_2\right)^2}_{\text{agent } (l=2)}. \tag{4.5}$$

Maximizing firm profits (4.5) over all players' effort levels yields that all players optimally exert the effort level $e_l^{FB} = \frac{p}{\theta}$. From this it follows directly that the first-best effort norm for employees is given by $n_l^{FB} = \frac{p}{\theta}$. First-best profits become:

$$\Pi^{FB} = \underbrace{\frac{1}{2}\frac{p^2}{\theta}}_{\text{principal } (l=0)} + \underbrace{\frac{1}{2}\frac{p^2}{\theta}}_{\text{supervisor } (l=1)} + \underbrace{\frac{1}{2}\frac{p^2}{\theta}}_{\text{agent } (l=2)}. \tag{4.6}$$

Lemma 4.1 *In the first-best case where effort is contractible, all players exert the effort level* $e_l^{FB} = \frac{p}{\theta}$. *The first-best effort norm for both employees therefore equals* $n_l^{FB} = \frac{p}{\theta}$.

Note that, when choosing her effort, the principal has two objectives in mind. On the one hand, she wants to maximize her individual production (first term of (4.5)). On the other hand, she wants to set the most efficient norm for the supervisor (second term of (4.5)). However, the principal does not face a trade-off between these two objectives. That is, the first-best effort level e_0^{FB} both maximizes the principal's individual production, and sets the most efficient norm for the supervisor. The reason is that the principal and the supervisor have the same marginal productivity of effort p. Moreover, even if the supervisor had a lower marginal productivity than the principal ($p_1 < p_0$), the principal would not face a trade-off between the objectives mentioned above. This holds because of the assumption that, to the extent that the principal's marginal productivity is higher, the supervisor does not consider

the principal's effort level to be normative (see equation (4.2)). Also note that the exact same considerations apply when the principal mandates the supervisor's effort level. In the next subsection, I show that when effort is not contractible, the principal does face a trade-off between maximizing the value of her own production and setting efficient norms for her employees.

4.4.2 Non-contractible effort

In case effort is not contractible, the first step in solving the model is to derive the employees' effort level for any given effort norm and compensation contract. Since the employee's wage w_l is fixed, the effort level that maximizes employee utility (4.1) equals:

$$e_l^* = \mu n_l, \tag{4.7}$$

where $\mu \equiv \frac{\gamma}{\gamma+\theta}$, and it holds that $0 < \mu < 1$. It can easily be verified that employee effort increases in μ, and therefore in the employee's norm sensitivity γ. The intuition is that deviating from the norm is more costly for an employee who possesses stronger norm sensitivity. Next, employee effort also increases in the norm n_l, as a higher norm implies higher costs of norm violation for a given level of effort. Finally, since it holds that $\mu < 1$, an employee always exerts less effort than his norm prescribes. The intuition is that employees trade off some costs of performing below the norm against the benefit of shirking. This simple fact has an important implication for hierarchical firms, namely that each employee transmits a norm towards the next layer of the organization that is lower than the norm he faces himself. Therefore, as one moves down the hierarchy, effort norms decrease. I call this phenomenon 'norm erosion'. The existence of norm erosion is the first result of the chapter, and all remaining results follow from this one.

Proposition 4.1 *A feature of hierarchical firms is norm erosion, that is,* $n_l^* > n_{l+1}^*$ *for all* l.

The second step in solving the model is to derive the optimal wage w_l. The optimal wage is the lowest possible wage that induces the employee to accept the

job, conditional on the effort level the employee exerts (e_l^*), and the norm he faces (n_l). This constraint reads $U_l\left(w_l, e_l^*, n_l\right) \geq 0$. Rewriting this condition to w_l yields:

$$w_l^* = \frac{1}{2}\theta\mu n_l^2. \tag{4.8}$$

The employee's wage increases in the norm he faces. The reason is two-fold. First, a higher norm induces the agent to exert more effort. Second, given his effort choice, a higher norm implies that the agent will incur higher costs of norm violation. The optimal wage also increases in μ, and therefore in employee norm sensitivity γ. Norm sensitivity has two effects on the wage. First, the employee incurs higher costs of effort if norm sensitivity is stronger, implying the wage must increase. Second, norm sensitivity has two effects on the employee's costs of norm violation. On the one hand, given his effort level, stronger norm sensitivity implies that the employee incurs greater psychological costs from performing below the norm. On the other hand, higher norm sensitivity implies that the employee exerts a higher effort level, which reduces the violation of the norm. However, the net effect of norm sensitivity on the wage is always positive, as made apparent by equation (4.8).

The final step in solving the model is to derive the principal's effort level. Conditional on the definition of effort norms (4.3), the employees' effort levels (4.7) and wages (4.8), total firm profits can be written as:

$$\Pi = \underbrace{pe_0 - \frac{1}{2}\theta e_0^2}_{\text{principal } (l=0)} + \underbrace{p\mu e_0 - \frac{1}{2}\theta\mu e_0^2}_{\text{supervisor } (l=1)} + \underbrace{p\mu^2 e_0 - \frac{1}{2}\theta\mu^3 e_0^2}_{\text{agent } (l=2)}. \tag{4.9}$$

Clearly, the principal's effort not only affects the value of her individual production (first term), but, by setting an example, also affects the value of hiring the supervisor (second term) and the agent (third term). This is also reflected in the effort level the principal optimally exerts. Maximizing firm profits over e_0 yields that this is equal to:

$$e_0^* = \frac{p}{\theta}\frac{1+\mu+\mu^2}{1+\mu+\mu^3}. \tag{4.10}$$

It can be easily checked that the principal's effort exceeds the first-best level derived in Lemma 4.1. This is the second result of the chapter:

Proposition 4.2 *The principal exerts an effort level that is higher than her first-best effort level ($e_0^* > e_0^{FB}$).*

To understand the result of Proposition 4.2, recall from Lemma 4.1 that the first-best norm for the supervisor and agent is given by $n_l^{FB} = \frac{p}{\theta}$. Hence, if the principal would exert the effort level $e_0^{FB} = \frac{p}{\theta}$, she would maximize the value of her individual production, and set the first-best norm for the supervisor. However, as derived in Proposition 4.1, the supervisor will erode the principal's example. The implication is that, if $e_0 = \frac{p}{\theta}$, the equilibrium norm for the agent will be inefficiently low. In response to this, the principal optimally sets a higher example to begin with, and thus raises her effort above e_0^{FB}.

Note that the result of Proposition 4.1 directly implies that the supervisor's norm is above the first-best level (recall that $n_l^{FB} = \frac{p}{\theta}$). In addition to this, it holds that the agent's norm is below the first-best level.[3] Intuitively, it can never be optimal that the agent's norm exceeds the first-best level. After all, driving the agent's norm closer to the first-best level imposes two costs on the principal: the principal's effort is inefficiently high (Proposition 4.2), and as a result, the supervisor's norm is inefficiently high. It follows that it may be optimal to raise the agent's norm up to the first-best level, but not higher than that. Hence, norm erosion ultimately implies the following Corollary.

Corollary 4.1 *Compared to the first-best benchmark, the effort norm for the supervisor is inefficiently high ($n_1^* > n_1^{FB}$), whereas the effort norm for the agent is inefficiently low ($n_2^* < n_2^{FB}$).*

Another result that norm erosion gives rise to is described in Corollary 4.2.

Corollary 4.2 *Firms optimally keep the number of organizational layers to a minimum.*

The intuition behind Corollary 4.2 is straightforward. So far, the analysis has shown that hierarchical firms are confronted with three inefficiencies in creating

[3] To see this formally, note that the inequality $n_2^* < n_2^{FB}$ reduces to $\frac{1-\mu^2}{1+\mu+\mu^3} > 0$, which always holds because $0 < \mu < 1$.

optimal norms for employees: the principal works too hard, the supervisor's norm is too high, and the agent's norm is too low. Hierarchical firms incur these costs because of norm erosion, which arises if intermediate layers of management are in place. Moreover, for any organizational layer added to the firm, the inefficiencies from norm erosion would be magnified. It follows that firms optimally keep the number of organizational layers to a minimum. In the remainder of the chapter, I assume that the principal requires supervisors to let agents do their job. A motivation for this assumption is that the principal simply cannot supervise all the agents she hires, and hence needs to delegate this task. Also see subsection 4.5.2, where I study how many supervisors and agents the principal optimally hires.

4.5 Implications for organizational design

This section offers two extensions of the basic model analyzed above. Both extensions yield a result as to how hierarchical firms can be better designed to reduce the inefficiencies stemming from norm erosion. In following order, I discuss promotion decisions and managerial spans of control.

4.5.1 Promotion decisions

In contrast to what has been assumed in the basic model, employees are likely to differ in their sensitivity to social norms. Given this heterogeneity, a natural question is which kind of employee the principal should optimally promote to the position of supervisor. To study this question, denote by γ_q the norm sensitivity of employee q, and by γ_r the norm sensitivity of employee r. Using the definition $\mu \equiv \frac{\gamma}{\theta+\gamma}$, one can also define μ_q and μ_r. Firm profits can now be written as:

$$\Pi = \underbrace{pe_0 - \frac{1}{2}\theta e_0^2}_{\text{principal } (l=0)} + \underbrace{p\mu_q e_0 - \frac{1}{2}\theta\mu_q e_0^2}_{\text{supervisor } (l=1)} + \underbrace{p\mu_q\mu_r e_0 - \frac{1}{2}\theta\mu_q^2\mu_r e_0^2}_{\text{agent } (l=2)}, \tag{4.11}$$

where it has been assumed that employee q is appointed supervisor.

Next, the principal optimally exerts the level of effort that maximizes firm profits.

This effort level can be shown to be equal to:

$$e_0^* = \frac{p}{\theta}\frac{1+\mu_q+\mu_q\mu_r}{1+\mu_q+\mu_q^2\mu_r}, \tag{4.12}$$

which is greater than the first-best level, as in Proposition 4.2. Equilibrium profits become:

$$\Pi^* = \frac{1}{2}\frac{p^2}{\theta}\frac{\left(1+\mu_q+\mu_q\mu_r\right)^2}{1+\mu_q+\mu_q^2\mu_r}. \tag{4.13}$$

Note that in case employee r is promoted to the position of supervisor, equilibrium profits are simply given by interchanging the subscripts q and r in (4.13).

To determine whether the principal optimally promotes employee q or r, one can compute which case yields higher profits. Promoting employee q yields higher profits than promoting employee r if:

$$\frac{1}{2}\frac{p^2}{\theta}\frac{\left(1+\mu_q+\mu_q\mu_r\right)^2}{1+\mu_q+\mu_q^2\mu_r} > \frac{1}{2}\frac{p^2}{\theta}\frac{\left(1+\mu_r+\mu_r\mu_q\right)^2}{1+\mu_r+\mu_r^2\mu_q}. \tag{4.14}$$

After some rewriting, this inequality reduces to $\mu_q > \mu_r$. In words, it is most profitable to promote the employee who has the strongest sensitivity to norms.

Proposition 4.3 *Hierarchical firms optimally promote employees with the strongest sensitivity to social norms.*

The intuition behind this result can be explained as follows. If the supervisor possesses stronger norm sensitivity, norms are eroded to a lower extent. Less norm erosion is always valuable to the principal. On the one hand, for any given effort choice of the principal, the agent will face a higher norm in equilibrium. Since the agent's norm is always inefficiently low (Corollary 4.1), this is profit-increasing. On the other hand, the principal can reduce her effort such that the agent's norm remains the same. This is also profit-increasing, since the principal's effort and the supervisor's norm are always inefficiently high (Proposition 4.2 and Corollary 4.1).

4.5.2 Managerial spans of control

The basic model assumes that the principal hires only one supervisor and one agent. More realistic is that the firm hires a number of S supervisors and A agents. Moreover, firms tend to maintain some relationship between the number of hired supervisors and agents. In the literature, the number of agents heading under one supervisor is referred to as the supervisor's span of control. Analogously, the number of supervisors heading under the principal is referred to as the principal's span of control. In this extension, I study how managerial spans of control are affected by the principal's desire to create efficient norms for her employees.

In case the firm hires S supervisors and A agents, firm profits are given by:

$$\Pi = \underbrace{pe_0 - \frac{1}{2}\theta e_0^2}_{\text{principal } (l=0)} + \underbrace{S\left(pe_1 - w_1\right) - \frac{1}{2}k_1 S^2}_{\text{supervisors } (l=1)} + \underbrace{A\left(pe_2 - w_2\right) - \frac{1}{2}k_2 A^2}_{\text{agents } (l=2)}, \tag{4.15}$$

where w_1 and w_2 must satisfy the employees' participation constraint (4.4). The profit function above is different from the one used in the basic model in two respects. First, to obtain an interior solution, I assume that the marginal profitability from hiring a supervisor or agent is decreasing. This is reflected by the terms $-\frac{1}{2}k_1 S^2$ and $-\frac{1}{2}k_2 A^2$. Second, the parameters k_1 and k_2 together describe the supervision technology of the firm. As will be shown below, for low values of k_1 and high values of k_2, the principal needs relatively many supervisors to supervise the agents, whereas for high values of k_1 and low values of k_2 the reverse holds.

Before I derive the optimal managerial spans of control, it is instructive to consider the first-best benchmark where effort is contractible. Using the results from Lemma 4.1, it is straightforward to derive the following Lemma.

Lemma 4.2 *In the first-best case where effort is contractible, the principal hires* $S^{FB} = \frac{1}{2}\frac{p^2}{\theta k_1}$ *supervisors and* $A^{FB} = \frac{1}{2}\frac{p^2}{\theta k_2}$ *agents. The first-best supervisor span of control is equal to* $\left(\frac{A}{S}\right)^{FB} = \frac{k_1}{k_2}$.

In the first-best case, there is no moral hazard problem in effort provision. Since supervisors and agents have the same marginal productivity of effort, it follows that differences in the marginal profitability of hiring a supervisor or agents can only

arise because of differences in the technology parameters k_1 and k_2. As a result, in the first-best case, the optimal supervisor span of control is uniquely determined by k_1 and k_2.

In the remainder of this subsection, I study the optimal managerial spans of control in case effort is not contractible. Given the definition of effort norms (4.3), employees' effort choice (4.7), and the employees' salary (4.8), firm profits can be written as:

$$\Pi = \underbrace{pe_0 - \frac{1}{2}\theta e_0^2}_{\text{principal } (l=0)} + \underbrace{S\left(p\mu e_0 - \frac{1}{2}\theta\mu e_0^2\right) - \frac{1}{2}k_1 S^2}_{\text{supervisors } (l=1)} + \underbrace{A\left(p\mu^2 e_0 - \frac{1}{2}\theta\mu^3 e_0^2\right) - \frac{1}{2}k_2 A^2}_{\text{agents } (l=2)}. \tag{4.16}$$

The principal will exert the effort level that maximizes (4.16), which is given by:

$$e_0^* = \frac{p}{\theta}\frac{1 + S\mu + A\mu^2}{1 + S\mu + A\mu^3}. \tag{4.17}$$

It holds that $e_0^* > e_0^{FB}$ for any $A > 0$, as in Proposition 4.2. It can easily be checked that the principal's effort decreases in S, but increases in A.[4] To understand these results, remember that the norm for supervisors is always above its first-best level, whereas the norm for agents is always below its first-best level (Corollary 4.1). Further, if the number of supervisors increases, their organizational layer makes up a larger share of firm profits. Therefore it becomes more important for the principal to have supervisors face an efficient norm. It follows that the principal optimally reduces her effort if S increases. The reverse intuition holds for the number of agents hired. The principal thus adjusts her effort in the direction that improves the effort norm for the kind of employee that becomes more abundant.

Proposition 4.4 *The principal's effort decreases in the number of supervisors hired,*

[4] To see this, note that:

$$\begin{aligned}\frac{\partial e_0^*}{\partial S} &= -\frac{p}{\theta}\frac{\mu^3 A(1-\mu)}{(1 + S\mu + A\mu^3)^2} < 0, \text{ and} \\ \frac{\partial e_0^*}{\partial A} &= \frac{p}{\theta}\frac{\mu^2(1-\mu)(1+S\mu)}{(1 + S\mu + A\mu^3)^2} > 0.\end{aligned}$$

$\frac{\partial e_0^*}{\partial S} < 0$, *and increases in the number of agents hired,* $\frac{\partial e_0^*}{\partial A} > 0$.

The next step is to determine the optimal values of S and A. Given the principal's effort choice (4.17), firm profits can be written as:

$$\Pi = \frac{1}{2}\frac{p^2}{\theta}\frac{(1+S\mu+A\mu^2)^2}{1+S\mu+A\mu^3} - \frac{1}{2}k_1S^2 - \frac{1}{2}k_2A^2. \tag{4.18}$$

The first-order conditions to S and A are given by:

$$\frac{1}{2}\frac{p^2}{\theta}\mu\frac{1+A\mu^4(2S-A+2A\mu)+2A\mu^3+S^2\mu^2+2S\mu}{(1+S\mu+A\mu^3)^2} = k_1S, \text{ and} \tag{4.19}$$

$$\frac{1}{2}\frac{p^2}{\theta}\mu^2\frac{(2-\mu)(1+2S\mu+S^2\mu^2)+A\mu^2(2+2S\mu+A\mu^3)}{(1+S\mu+A\mu^3)^2} = k_2A, \tag{4.20}$$

respectively. Although the first-order conditions are difficult to solve algebraically, they are instructive in two ways.

First, given the first-best solutions S^{FB} and A^{FB} derived in Lemma 4.2, both (4.19) and (4.20) are negative.[5] Hence, the first-best solutions cannot be optimal in case effort is not contractible. Moreover, since k_1 and k_2 are assumed to be sufficiently large such that the second-order conditions are negative, in the optimum it must hold that $S^* < S^{FB}$ and $A^* < A^{FB}$.[6] The intuition behind this result is straightforward. In case effort is not contractible, a moral hazard problem in effort

[5] One can verify that, given the first-best solutions from Lemma 4.2, (4.19) and (4.20) reduce to:

$$\frac{\partial\Pi}{\partial S} = -\frac{\frac{1}{2}\frac{p^2}{\theta}(1-\mu)}{(k_1p^2\mu^3+k_2p^2\mu+2\theta k_1k_2)^2}\left(\begin{array}{l}p^4\mu^5k_1^2+2p^4\mu^4k_1k_2+p^4\mu^2k_2^2+\\ 4p^2\theta\mu^3k_1^2k_2+4p^2\theta\mu k_1k_2^2+4\theta^2k_1^2k_2^2\end{array}\right) < 0, \text{ and}$$

$$\frac{\partial\Pi}{\partial A} = -\frac{\frac{1}{2}\frac{p^2}{\theta}(1-\mu)}{(k_1p^2\mu^3+k_2p^2\mu+2\theta k_1k_2)^2}\left[\begin{array}{l}p^2\mu^3k_1\left(2\mu p^2k_2+\mu^3p^2k_1+4\theta k_1k_2\right)+\\ 4\theta\mu\left(p^2k_1k_2^2+\theta k_1^2k_2^2\right)\left(1+\mu-\mu^2\right)+\\ p^4\mu^2k_2^2\left(1-\mu^2+\mu^3\right)\end{array}\right] < 0,$$

where the signs follow from $0<\mu<1$, $p>0$, $\theta>0$, $k_1>0$, and $k_2>0$.

[6] The second-order conditions are given by:

$$\frac{\partial^2\Pi}{\partial S^2} = \frac{1}{2}\frac{p^2}{\theta}2\mu^6(1-\mu)^2\frac{A^2}{(A\mu^3+S\mu+1)^3} - k_1, \text{ and}$$

$$\frac{\partial^2\Pi}{\partial A^2} = \frac{1}{2}\frac{p^2}{\theta}2\mu^4(1-\mu)^2\frac{(S\mu+1)^2}{(A\mu^3+S\mu+1)^3} - k_2,$$

which are always negative if k_1 and k_2 are sufficiently large.

provision arises. This problem is partly mitigated because employees are sensitive to social norms. However, as shown in Corollary 4.1, equilibrium norms for supervisors are always above first-best, whereas equilibrium norms for agents are always below first-best. As a result, it is less attractive to hire supervisors and agents. The implication is that the principal's span of control is distorted downwards relative to the first-best case.

Second, the marginal benefit of hiring an agent (LHS of (4.20)) may be higher than the marginal benefit of hiring a supervisor (LHS of (4.19)). Rewriting this inequality yields:

$$A > \frac{1 + S\mu}{\mu^2}. \tag{4.21}$$

Hence, given that agents are relatively abundant, the marginal benefit of hiring an agent exceeds the marginal benefit from hiring a supervisor. The intuition behind this result stems directly from Proposition 4.4. Recall that this Proposition essentially states that the principal sets her effort level such that the most abundant kind of employee faces the more efficient norm in equilibrium. Therefore, given that agents are relatively abundant, the principal ensures that the equilibrium norm for agents is closer to the first-best level than the equilibrium norm for supervisors.[7] It follows that hiring agents yields a higher marginal benefit. Of course, the reverse may also hold: given that supervisors are relatively abundant, their marginal benefit is higher.

Finally, each of the mechanisms described above may be at work in equilibrium, depending on the values of k_1 and k_2. Figures 4.1 and 4.2 provide contour plots of the profit function (4.18) for a certain parameterization of the model. The inner contours represent a higher profit level than the outer contours, implying that the maximum of the profit function is represented by the dots in the inner contours. In both figures, the 45-degrees line depicts the first-best supervisor span of control, $\left(\frac{A}{S}\right)^{FB} = \frac{k_1}{k_2}$. Figure 4.1 depicts a case where the first-best supervisor span of control is low ($\frac{k_1}{k_2} = 0.1$). Next, as can be seen in the plot, the maximum of firm profits (4.18) lies above the 45-degrees line, meaning that the principal distorts the

[7] In fact, the norm for agents is closer to the first-best level than the norm for supervisors when it holds that $\left|n_2^* - n_2^{FB}\right| < \left|n_1^* - n_1^{FB}\right|$. Taking into account that $n_1^* = e_0^*$, $n_2^* = \mu e_0^*$, and $n_1^{FB} = n_2^{FB} = \frac{p}{\theta}$, one can show that this inequality is identical to condition (4.21).

supervisor's span of control downwards relative to the first-best case. The reason is that the supervision technology is such that it is attractive to hire relatively many supervisors. As described above, given that supervisors are relatively abundant, the principal optimally makes sure that supervisors' effort norms are closer to the first-best level than agents' effort norms. As a result, supervisors become even more attractive to hire, implying that the principal optimally distorts the supervisor's span of control downwards.

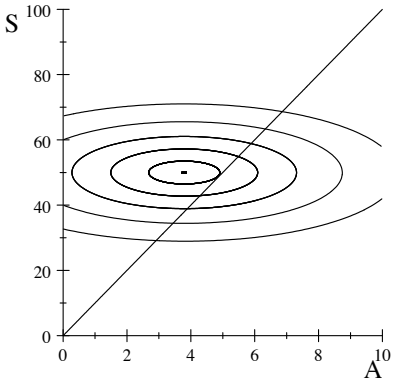

Figure 4.1. Firm profits as a function of S and A. $\frac{p^2}{\theta} = 20$, $\mu = 0.5$, $k_1 = 0.1$, and $k_2 = 1$.

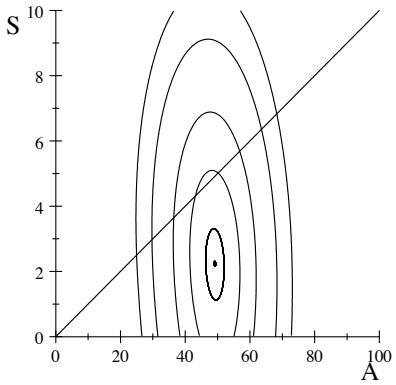

Figure 4.2. Firm profits as a function of S and A. $\frac{p^2}{\theta} = 20$, $\mu = 0.5$, $k_1 = 1$, and $k_2 = 0.1$.

Figure 4.2 depicts a case where the first-best supervisor span of control is high ($\frac{k_1}{k_2} = 10$). By the same logic as above, the principal then optimally distorts the supervisor's span of control upwards. The final Proposition summarizes the results of this subsection.

Proposition 4.5 *Relative to the first-best case, hierarchical firms optimally distort the principal's span of control downwards. Relative to the first-best case, the supervisor's span of control is optimally distorted downwards if* $\frac{k_1}{k_2}$ *low, and optimally distorted upwards if* $\frac{k_1}{k_2}$ *is high.*

4.6 Concluding remarks

This chapter sets out to study the problem of creating effort norms within hierarchical firms. The analysis presented above rests on one key premise, namely that the actual example of superiors determines effort norms for subordinates. Norms are therefore shaped by two factors: the ultimate example set by those at the top of the hierarchy, like the CEO, and the process of transmitting this example to lower levels of the firm. Importantly, firms cannot control the process of norm transmission. After all, each hierarchical layer sets her own example for the next one. Consequently, norm transmission is prone to a moral hazard problem: no employee fully conforms to his norm, as conforming requires him to exert costly effort. The implication is that, in equilibrium, norms erode as one moves down the hierarchy. To counteract the norm erosion effect, it has been shown that top managers optimally set a higher example to begin with. This is well in line with the fact that top managers work exceptionally long and hard. In addition to this, norm erosion gives rise to two comparative static results on the effort of top managers. Their effort decreases in the number of middle-managers hired, and increases in the number of lower-level employees hired.

The analysis also yields three implications for organizational design. First, the inefficiencies due to norm erosion imply that firms optimally keep the extent of hierarchy to a minimum. Second, to reduce norm erosion, hierarchical firms optimally promote employees with the strongest sensitivity to social norms. Last, hierarchical firms optimally distort supervisor spans of control. The mechanism driving this result is that the relative profitability of hiring middle-managers and lower-level employees depends on whose equilibrium norm is more efficient.

4.A Allowing for performance pay

In this Appendix, I show that the results of the chapter are not sensitive to the introduction of performance pay, as long as using performance pay generates some kind of agency costs.[8] Suppose that employee effort yields a verifiable signal $y_l = pe_l + \varepsilon$, where ε reflects noise in the signal. The expected value of ε equals $E[\varepsilon] = 0$, and ε has a variance equal to $\sigma_\varepsilon^2 > 0$. Denote by s_l the base salary and by b_l the incentive intensity of the employee's contract. I assume that using performance pay is costly because employees are risk-averse. To model this assumption, employee utility is given by:

$$U_l = -\exp^{-r\left[s_l + b_l(pe_l+\varepsilon) - \frac{1}{2}\theta e_l^2 - \frac{1}{2}\gamma(n_l - e_l)^2\right]}, \tag{4.22}$$

where $r > 0$ denotes the intensity of risk-aversion. The certainty equivalent of (4.22) is given by:

$$E[U_l] = s_l + b_l p e_l - \frac{1}{2}\theta e_l^2 - \frac{1}{2}\gamma(n_l - e_l)^2 - \frac{1}{2}r\sigma_\varepsilon^2 b_l^2. \tag{4.23}$$

This is the same utility function as in (4.1), but with two terms added. The term $b_l p e_l$ describes the incentive portion of the employee's compensation. The term $\frac{1}{2}r\sigma_\varepsilon^2 b_l^2$ describes the risk-related disutility which performance pay imposes upon the employee.

The model can now be solved, following the same three steps as in subsection 4.4.2. First, given the contract and effort norm, employees choose the effort level that maximizes their utility. This effort level can be shown to be equal to:

$$e_l^* = \lambda b_l + \mu n_l, \tag{4.24}$$

where $\lambda \equiv \frac{p}{\theta+\gamma}$ and $\mu \equiv \frac{\gamma}{\theta+\gamma}$. Equation (4.24) shows that employee effort not only increases in the norm, but also in the incentive intensity of his contract.

Second, given employee effort (4.24), the lowest possible base salary that ensures

[8] As will be shown below, if there are no agency costs of using incentives, the principal can achieve the first-best outcome derived in section 4.4.1.

participation can be written as:

$$s_l^* = \frac{1}{2}\theta\mu n_l^2 + \frac{1}{2}p\lambda b_l^2 + \frac{1}{2}r\sigma_\varepsilon^2 b_l^2 - b_l p e_l^*. \tag{4.25}$$

Last, given employee effort (4.24) and the base salary (4.25), the principal chooses her effort and the incentive intensities such that firm profits are maximized. This problem can be written as:

$$\begin{aligned}\max_{e_0,b_1,b_2}\Pi \;=\; & \underbrace{pe_0 - \frac{1}{2}\theta e_0^2}_{\text{principal } (l=0)} + \underbrace{p\left(\lambda b_1 + \mu e_0\right) - \frac{1}{2}\theta\mu e_0^2 - \frac{1}{2}p\lambda b_1^2 - \frac{1}{2}r\sigma_\varepsilon^2 b_1^2}_{\text{supervisor } (l=1)} + \\ & \underbrace{p\left[\lambda b_2 + \mu\left(\lambda b_1 + \mu e_0\right)\right] - \frac{1}{2}\theta\mu\left(\lambda b_1 + \mu e_0\right)^2 - \frac{1}{2}p\lambda b_2^2 - \frac{1}{2}r\sigma_\varepsilon^2 b_2^2}_{\text{agent } (l=2)}.\end{aligned} \tag{4.26}$$

The solutions to the maximization problem are given by:

$$e_0^* = \frac{p}{\theta}\frac{1+\mu+\mu^2-\theta\mu^2\lambda^2\frac{1+\mu}{r\sigma_\varepsilon^2+p\lambda+\theta\lambda^2\mu}}{1+\mu+\mu^3-\theta\mu^2\lambda^2\frac{\mu^2}{r\sigma_\varepsilon^2+p\lambda+\theta\lambda^2\mu}}, \tag{4.27}$$

$$b_1^* = \frac{p\lambda+2p\lambda\mu}{\theta\lambda^2\mu\left(1+\mu\right)+\left(p\lambda+r\sigma_\varepsilon^2\right)\left(1+\mu+\mu^3\right)}, \text{ and} \tag{4.28}$$

$$b_2^* = \frac{p\lambda}{p\lambda+r\sigma_\varepsilon^2}. \tag{4.29}$$

It is easy to check that, if using incentives is costly ($r > 0$ and $\sigma_\varepsilon^2 > 0$), it holds that $1 > b_1^* > b_2^*$. Hence, both employees receive an incentive share below 1, implying that employees will not fully internalize the benefits of their effort to the principal. As a result, norms will be eroded, as in Proposition 4.1. Since all remaining results of the chapter follow from this one, it holds that the results are insensitive to the introduction of performance pay. However, note that in case using incentives is costless ($r = 0$ or $\sigma_\varepsilon^2 = 0$), norm erosion can be avoided and all players exert the first-best effort level, $e_l^* = e_l^{FB} = \frac{p}{\theta}$. The reason is that the principal optimally provides full-powered incentives to the employees, that is, $b_1^* = b_2^* = 1$.

Finally, if using incentives is costly, it is found that the supervisor receives stronger incentives than the agent. The intuition behind this result is that pro-

viding incentives to the supervisor has the additional benefit of raising the norm for the agent. Interestingly, this result is in line with Aggarwal and Samwick (2003), who document that CEOs receive stronger incentives than executives with oversight authority, who in turn receive stronger incentives than managers with divisional responsibility.

Chapter 5

Summary

This thesis has studied the problem of motivating employees. From an economist's perspective, motivation problems arise because of two reasons. First, labor relationships are characterized by a divergence of interests: putting in hard work benefits the employer, whereas the costs of it are borne by the employee. It is therefore in the employee's interest to shirk. Second, it is usually impossible to force employees to work hard by means of a legal contract, because in most cases employees' effort cannot be verified by a judge. As a result, organizations must motivate their employees to act in the organization's interest (Laffont and Martimort 2002).

Economists traditionally emphasize the importance of monetary incentives for employee motivation. This is not without merit, since empirical evidence shows that employees indeed work harder when doing so increases their compensation (see e.g. Lazear 2000 and Shearer 2004; for overviews of the literature see Prendergast 1999 and Lazear and Oyer 2013). However, over the last decades economists have started to draw a broader picture of employee motivation. Among the alternative motivational factors that have been studied intensively are reciprocity (e.g. Akerlof 1982 and Non 2012), intrinsic motivation for specific tasks (e.g. Besley and Ghatak 2005 and Delfgaauw and Dur 2007), and inequity aversion (e.g. Fehr and Schmidt 1999 and Bolton and Ockenfels 2000). The present thesis contributes to this relatively new strand of literature. It studies how altruism and the preference to conform to social norms affect employee motivation, and derives implications for incentive contracts and organizational design. The research conducted here builds on

principal-agent models, and is theoretical in nature. The scientific value of this thesis therefore lies primarily in developing new hypotheses and interpreting observed facts. However, some of the results also have clear implications for management. In the remainder of this chapter, I summarize the key features of the models developed, and the results derived from analyzing them.

Chapters 2 and 3 of this thesis have two common denominators. First, both chapters study how altruism and spite between an agent and a principal or supervisor affect optimal incentive contracts. By 'altruism' I mean the preference for someone else to be better off. The opposing concept 'spite' refers to the preference for someone else to be worse off. Altruism and spite are therefore formally modeled as the dependence of a player's utility on some other player's utility. Second, both chapters study the provision of incentives in situations where the agent's performance is subjectively evaluated, which is widely observed in practice (see e.g. Gibbons 1998, Prendergast 1999, Gibbs et al. 2004 and Gibbs 2012). If performance is subjectively evaluated, the principal cannot write explicit contracts on the agent's performance, since a subjective evaluation cannot be verified by a judge. Two alternative ways to incentivize the agent are studied in this thesis: incentives based on a relational, or implicit, contract (chapter 2), and delegation of performance evaluation to a non-residual claimant supervisor (chapter 3).

In chapter 2, Robert Dur and I study how relational incentive contracts are affected in case a principal and agent may be altruistic or spiteful toward each other. A relational incentive contract motivates the agent by means of promises and/or threats based on the principal's subjective evaluation of the agent's performance. Because of the subjectivity involved, a relational contract cannot be enforced in court. To be effective, a relational contract must therefore be credible. That is, the principal must be willing to actually honor her promises and carry out her threats. This is only the case if the following holds: the benefits from deviating from the contract now, are outweighed by the costs of being punished for the deviation in the future. To allow for this condition to be satisfied in equilibrium, we make two standard assumptions. First, the game is infinitely repeated, and second, the agent plays a trigger strategy that prescribes to always shirk after the principal

breaks a relational contract. Note that the credibility requirement described above may considerably constrain the principal's promises and threats, and therefore the incentive power of relational contracts.

In the model developed in chapter 2, we allow the principal to use two means to motivate the agent: she may use a promise to pay a bonus in case the agent's performance is good, and she may use a threat to dismiss the agent in case his performance is bad. We also assume that both players have some bargaining power, which implies that the players divide the total value of their labor relationship between each other. Our first key finding is that principal's and agent's altruism improves the credibility of a promise to pay a bonus. The reason is that an altruistic principal finds it less costly to pay out the bonus. In addition to this, altruism makes the labor relationship more valuable since both players enjoy, to some extent, each others' utility. The implication is that it becomes less attractive for the principal to break her promise, as doing so puts the labor relationship at risk. Our second key finding is that principal's and agent's altruism may be too high for a threat of dismissal to be credible. The intuition is that, if altruism is high, the labor relationship is sufficiently valuable for the principal to hire the agent, even if the agent shirks. Hence, altruism undermines the credibility of a threat of dismissal.

The key results described above imply two counterintuitive comparative statics. First, even though a more altruistic agent requires lower incentives to be willing to exert effort, higher altruism may lead to higher bonuses. The intuition is that altruism makes a promise to pay a bonus credible in the first place. This result is in contrast with the existing theoretical literature, which predicts that labor relationships characterized by higher altruism have weaker incentives (e.g. Casadesus-Masanell 2004, Sliwka 2007, Shchetinin 2010, and Non 2012). Second, even though altruism implies that each player partly enjoys the other player's utility, a marginal increase in altruism may make players worse off in equilibrium. The reason is that a marginal increase in altruism may render a threat of dismissal not credible. If this is the case, the agent faces lower incentives to exert effort, which reduces the production value generated within the labor relationship.

Whereas chapter 2 studies the case of altruism and spite between a principal and

agent, in chapter 3 I focus on the case where a non-residual claimant supervisor may be altruistic toward the agent. In the model developed, it is the supervisor's task to subjectively evaluate the agent's performance and report this to the principal. The principal uses the supervisor's report to incentivize the agent. In case the supervisor reports the agent's performance to be high, the principal pays the agent a bonus. The supervisor may be one of two types: selfish, or altruistic toward the agent. I assume that the agent observes the supervisor's type, but the principal does not. As in chapter 2, the game is infinitely repeated, and the agent plays a trigger strategy. The trigger strategy prescribes to always shirk after the supervisor gives a biased performance evaluation. There is a large pool of supervisors and agents. Each period, the principal can offer a contract to one supervisor and one agent. Finally, I assume that the contract offered to the supervisor is uniform, that is, the principal cannot offer a menu of contracts to the supervisor.

In the model, the supervisor's altruism may affect the agent's incentives to exert effort. The reason is that an altruistic supervisor enjoys the well-being of the agent, and may therefore report high performance irrespective of the agent's actual performance. However, if the agent expects this, he has no reason to exert effort. The principal must therefore make sure that the supervisor will report performance truthfully. In chapter 3, I study how the principal achieves this through offering optimal incentive contracts to the supervisor and the agent. The analysis yields two key results.

First, in contrast to the existing literature (e.g. Prendergast and Topel 1996 and Giebe and Gürtler 2012), I show that supervisor's altruism may provide a net incentive to report performance truthfully, rather than to bias evaluations. As in the models mentioned above, supervisor's altruism increases the benefit from biasing the performance evaluation upwards. However, in contrast to these models, altruism also increases the benefits from reporting truthfully. The reason is that, because of his altruism, the altruistic supervisor enjoys a rent from working with the agent. This rent cannot be extracted from the supervisor by lowering his compensation, since doing so would imply that the selfish type is no longer willing to accept the contract. Next, I show that the supervisor's contract optimally makes sure that the

supervisor's utility is no higher than his reservation utility after rating the agent's performance incorrectly. Hence, biasing the performance evaluation implies that the altruistic supervisor loses the rent from working with the agent in the future. If the supervisor is sufficiently patient, the latter effect outweighs the former, implying that supervisor's altruism provides a net incentive to report performance truthfully.

Second, it may be optimal to design the supervisor's contract such that it attracts only one supervisor type. The crucial feature of such a screening contract is that it reveals the supervisor's type, conditional on the contract being accepted. Given that the principal knows the supervisor's type, she can eliminate the supervisor's benefit from biasing the performance evaluation. I show that the principal can achieve this by tying the supervisor's compensation to his performance report, as in Giebe and Gürtler (2012). The benefit to the principal is that the supervisor does not have to earn costly rents to be willing to report performance truthfully. For this reason, I show that screening for one supervisor type may be optimal, even though this reduces the probability that vacancies are filled.

Chapter 4 of the thesis studies another motivation in the workplace, namely the preference to conform to social norms. Given that employees conform to social norms, an important question is how firms create those norms that induce profit-maximizing behavior from their employees. The analysis in chapter 4 provides answers to this question. The innovation of the study compared to the existing literature (e.g. Sliwka 2007 and Fischer and Huddart 2008), is its assumption that norms within firms are affected by the actual behavior of superiors. I explore the implications of this assumption for norm creation in hierarchical firms. To this end, I develop a three-layer principal-supervisor-agent model. In the model, each player exerts effort which generates value for the principal's firm. Both employees (the supervisor and the agent) incur a psychological cost if their effort level falls below the norm they face. Each employee's effort norm is given by the actual effort level of his superior. Hence, the principal's example constitutes the norm for the supervisor, whose example in turn constitutes the norm for the agent.

The key result derived in chapter 4 is that hierarchical firms are confronted with 'norm erosion'. By norm erosion I mean that norms become lower as one moves

down the hierarchy. The mechanism driving this result is simple. When choosing how much effort to exert, each employee takes into account the norm set by his superior, but never fully complies with it. The reason is that employees find it privately optimal to trade off some costs of performing below the norm against the benefit of some shirking. As a result, each employee sets a lower example for his subordinate than the example he got from his superior.

The implication of the supervisor eroding the principal's example is that the agent faces an inefficiently low norm. I show that the principal takes this into account when choosing her own effort level, and optimally sets a higher example to begin with. This strategy is costly for two reasons, however. First, by setting such a high example, the principal incurs relatively high costs of effort. Second, the principal's high example imposes high effort costs and norm violation costs upon the supervisor, for which he must be compensated through his wage. Because of norm erosion, hierarchical firms thus incur three inefficiencies in norm creation: the principal's effort is inefficiently high, the supervisor's norm is inefficiently high, and the agent's norm is inefficiently low.

The final part of chapter 4 studies how hierarchical firms can be optimally designed so as to minimize the inefficiencies in norm creation. A straightforward implication of the analysis is that firms should keep the extent of hierarchy to a minimum, as this reduces the possibilities for norms to be eroded. Second, firms optimally promote employees with the strongest sensitivity to social norms. Supervisors with a stronger sensitivity to social norms conform to the principal's example to a larger extent, implying that her example is eroded to a lower extent. Last, firms optimally distort the supervisor's span of control, that is, the number of agents heading under one supervisor. The intuition is as follows. In the model, I assume that there is some exogenously given supervision technology which affects the optimal supervisor span of control. However, when choosing how many supervisors and agents to hire, the principal also takes into account the social norms that will arise in equilibrium. Recall that the supervisor's norm is always inefficiently high, whereas the agent's norm is always inefficiently low. Next, it holds that the magnitude of the inefficiency may differ for agents and for supervisors. In fact, the principal optimally makes sure

that the inefficiency is relatively small for the kind of employee that is relatively abundant. She achieves this by decreasing her effort level in the number of supervisors, and increasing her effort level in the number of agents. It follows that if the supervision technology favors a high (low) supervisor span of control, agents (supervisors) face a relatively efficient norm, which makes agents (supervisors) relatively more attractive to hire. As a result, firms optimally distort supervisors' spans of control away from its first-best level.

Samenvatting
(Summary in Dutch)

Motivatie, doel en methode

Een noodzakelijke voorwaarde voor het succes van een organisatie is dat haar werknemers ijverig hun taken uitvoeren. Echter, het is de vraag of werknemers zich zo ijverig gedragen, en wel om twee redenen. Ten eerste worden arbeidsrelaties gekarakteriseerd door uiteenlopende belangen: hard werken is waardevol voor de werkgever, maar de moeite die dat kost moet worden opgebracht door de werknemer. Het is daarom in het belang van de werknemer om wat rustiger aan te doen. Ten tweede is het veelal onmogelijk om door middel van een juridisch contract werknemers te dwingen hard te werken, omdat dit voor een rechter moeilijk te verifiëren is. Deze premissen, die de basis vormen van de principaal-agent theorie, impliceren dat organisaties hun werknemers moeten motiveren om goed hun best te doen (zie bijv. Laffont en Martimort 2002).

Traditioneel leggen economen veel nadruk op financiële prikkels om werknemers te motiveren. Dit is niet zonder reden. Goed opgezette studies laten zien dat werknemers hun prestaties verbeteren als zij hiermee hun salaris kunnen verhogen (zie bijv. Lazear 2000 en Shearer 2004; voor een overzicht van de literatuur zie Prendergast 1999 en Lazear en Oyer 2013). Toch is geld niet de enige motivatie op de werkvloer. Misschien wel de belangrijkste aanwijzing hiervoor is dat maar weinig werknemers een prestatieafhankelijke beloning krijgen (zie bijv. MacLeod en Parent 2000). Zelfs als dit wel het geval is, is het deel van het salaris dat van prestaties afhankelijk is gewoonlijk klein. Gedurende de laatste decennia hebben economen dan

ook veel studie gemaakt van niet-financiële motivaties op de werkvloer. Voorbeelden daarvan zijn wederkerigheid (bijv. Akerlof 1982 en Non 2012), intrinsieke motivatie voor bepaalde taken (bijv. Besley en Ghatak 2005 en Delfgaauw en Dur 2007), conformisme (bijv. Sliwka 2007), altruïsme (bijv. Prendergast en Topel 1996), en ongelijkheidsaversie (bijv. Fehr en Schmidt 1999 en Bolton en Ockenfels 2000).

Dit proefschrift draagt bij aan de economische literatuur over niet-financiële motivatie op de werkvloer. Het doel is om het belang van altruïsme en conformisme voor de motivatie van werknemers te onderzoeken. De hoofdstukken 2 en 3 bestuderen de implicaties van altruïsme tussen een werknemer en zijn werkgever of manager voor het optimaal gebruik van financiële prikkels. Beide hoofdstukken zijn gericht op situaties waar de prestaties van werknemers subjectief beoordeeld worden. In hoofdstuk 4 veronderstel ik dat werknemers gevoelig zijn voor normen betreffende hun inzet. Ik zoek vervolgens antwoorden op de vraag hoe organisaties optimaal zulke normen creëren, en wat dit betekent voor hoe organisaties optimaal ontworpen worden. Het onderzoek in dit proefschrift vindt plaats door middel van de ontwikkeling en analyse van principaal-agent modellen, en is theoretisch van aard. De wetenschappelijke waarde van dit proefschrift ligt daarom voornamelijk in het ontwikkelen van nieuwe hypothesen en de interpretatie van geobserveerde feiten. De resultaten van het onderzoek hebben echter ook praktische implicaties voor de werkvloer.

Het resterende deel van dit hoofdstuk is als volgt georganiseerd. In de volgende sectie bespreek ik de concepten altruïsme en conformisme, en motiveer het belang van beide voor de werkvloer. Daarna sluit ik af met een samenvatting van de modellen die in dit proefschrift ontwikkeld zijn, en de belangrijkste inzichten waartoe deze modellen leiden.

Altruïsme en conformisme op de werkvloer

De eerste gedragsassumptie die bestudeerd wordt in dit proefschrift is die van altruïsme. De Oxford English Dictionary definieert ‘altruism’ als een "disinterested and selfless concern for the well-being of others", ofwel, het geven om het welzijn van anderen zonder bijbedoelingen of zelfzucht. Er zijn veel aanwijzingen dat ten

minste sommige mensen altruïstisch zijn. Een goed voorbeeld vormen de talloze experimenten waarin mensen de dictator game spelen. In dit spel beslist een speler hoe een vaste som geld te verdelen tussen zichzelf en een andere speler. In tegenstelling tot de voorspelling van de klassiek economische theorie, waarin verondersteld wordt dat mensen hun geldelijk inkomen maximaliseren, geeft gewoonlijk 60% van de spelers geld aan de andere speler. De gemiddelde grootte van de transfer is 20% van het te verdelen geld (Levitt en List 2007). Een andere aanwijzing voor het bestaan van altruïsme vormt het geobserveerde gedrag in public good games. In dit type spel besluit elke speler hoeveel geld van zijn eigen vermogen bij te dragen aan een pot. De pot wordt verdeeld onder alle n spelers, nadat de waarde ervan vermenigvuldigd is met een factor groter dan 1. Vanwege deze vermeerdering is het optimaal voor de groep als iedere speler zijn totale vermogen bijdraagt aan de pot. Echter, als een speler alleen om zijn eigen geldelijke inkomen geeft, is het individueel optimaal om niets bij te dragen aan de pot. De reden is dat iedere speler de volledige kosten van zijn contributie draagt, maar slechts een fractie $1/n$ van de baten van zijn contributie geniet. Toch liggen de contributies gemiddeld rond de 40% van het groep optimum (Ledyard 1995), wat suggereert dat spelers om het welzijn van anderen geven. Experimenten zoals hierboven beschreven suggereren niet alleen dat sommige mensen altruïstisch zijn, maar ook dat sommige mensen afgunstig zijn. Andreoni en Miller (2002), bijvoorbeeld, vinden dat 55% van hun onderzoekssubjecten bereid is om geld op te geven om het vermogen van een ander te verhogen, maar dat 23% bereid is om geld op te geven om het vermogen van een ander te verlagen.

Het geobserveerde gedrag in dictator games en public good games is echter niet zonder meer een bewijs van het bestaan van altruïsme. Empirisch onderzoek laat zien dat niet al het gedrag dat altruïstisch lijkt, ook daadwerkelijk gedreven wordt door dit motief. Landry et al. (2006), bijvoorbeeld, bestuderen in een veld experiment donaties tijdens deur-aan-deur campagnes voor het goede doel. De auteurs laten zien dat de hoogte van donaties stijgt in de aantrekkelijkheid van de vrouwelijke collectanten. In een vergelijkbare setting bestuderen DellaVigna et al. (2012) het effect van een flyer aan de deurkruk die aangeeft wanneer een collectant langs komt. De flyer vermindert het aantal keren dat mensen de deur open doen significant, wat

consistent is met de hypothese dat mensen aan het goede doel geven vanwege sociale druk. Echter, DellaVigna et al. (2012) vinden dat het aandeel grote donaties (meer dan $10) niet wordt beïnvloedt door de flyer. De auteurs vatten dit op als indirect bewijs voor het bestaan van altruïsme.

De moeilijkheid in het debat over altruïsme is dat het onmogelijk is om het bestaan ervan te verifiëren. Men kan gedrag observeren, maar niet de motieven erachter. Dientengevolge kan men nooit uitsluiten dat schijnbaar altruïstisch gedrag door een ander motief wordt gedreven. Men heeft daarom geprobeerd om het bestaan van altruïsme te falsifiëren, door zoveel mogelijk alternatieve motivaties uit te sluiten. De stand van deze literatuur is dat het gedrag in, bijvoorbeeld, dictator games en public good games deels kan worden verklaard door de prikkel om een goede reputatie op te bouwen, het genot van het geven, wederkerigheid en sociale druk. Echter, de mogelijkheid van altruïsme kan door dit type onderzoek niet worden uitgesloten. Het geobserveerde gedrag in dictator games en public good games vormt daarom een sterke aanwijzing dat sommige mensen altruïstisch zijn (zie Andreoni et al. 2008 voor een overzicht van de literatuur).

Er is ook meer specifiek bewijs ten aanzien van het belang van altruïsme voor relaties tussen werknemers en hun werkgevers en managers. Enquêtes onder managers laten zien dat vriendschap tussen managers en werknemers regelmatig voorkomt (zie bijv. Berman et al. 2002). Campbell en Kamlani (1997) rapporteren dat een meerderheid van Amerikaanse managers vindt dat goede relaties tussen managers en werknemers belangrijker zijn voor de motivatie van werknemers dan goede arbeidsomstandigheden, hoge lonen en controle.[1] Er is ook bewijs voor het bestaan van slechte relaties tussen managers en werknemers. Moerbeek en Need (2003) rapporteren Nederlandse data waaruit blijkt dat in acht procent van de banen die respondenten hadden, ze slecht tot zeer slecht konden opschieten met hun manager. Iets suggestiever is de studie van Kahneman et al. (2004), die op basis van data uit dagboeken laat zien dat van alle reguliere activiteiten op een dag, Amerikaanse vrouwen de grootste hekel hadden aan interactie met hun baas.

[1] In overeenstemming hiermee documenteert een uitgebreide literatuur in de organisatie psychologie een sterk positieve correlatie tussen de kwaliteit van manager-werknemer relaties en de prestaties van de werknemers (zie bijv. Wayne et al. 1997 en Rhoades en Eisenberger 2002).

Altruïsme tussen werknemers en hun werkgever of manager heeft verschillende implicaties voor arbeidsrelaties. In het geval van werkgever-werknemer relaties, bijvoorbeeld, kan altruïsme motivatieproblemen verzachten. De reden is dat altruïstische werknemers een prikkel hebben om bij te dragen aan het welzijn van de werkgever. Gebaseerd op deze logica is er een theoretische literatuur die voorspelt dat altruïstische werknemers lagere financiële prikkels krijgen (bijv. Casadesus-Masanell 2004, Sliwka 2007, Shchetinin 2010, en Non 2012). Een ander voorbeeld is dat goede sociale relaties op de werkvloer een baan waardevoller maken voor medewerkers. Economische theorie voorspelt daarom dat goede sociale relaties op de werkvloer een substituut zijn voor hoge lonen (bijv. Grund en Sliwka 2005, Arce 2010, Dur en Sol 2010, en hoofdstukken 2 en 3 van dit proefschrift). Borzaga en Depedri (2005) rapporteren enige empirische ondersteuning voor deze predictie. Voor de Italiaanse non-profit sector vinden de auteurs een negatieve correlatie tussen de kwaliteit van sociale relaties en lonen.

De tweede gedragsassumptie die bestudeerd wordt in dit proefschrift is dat mensen hun gedrag willen conformeren aan sociale normen. Een vroeg, onder economen berucht experiment dat dit fenomeen documenteert is Asch (1951). In dit experiment kregen mensen de gemakkelijke taak om lijnen van gelijke lengte te matchen. Wanneer mensen deze taak individueel uitvoerden werd een gemiddelde accuratesse van 98% procent bereikt. De experimentele behandeling bestond in het uitvoeren van dezelfde taak, maar dan in een groep van acht mensen. Zonder medeweten van de experimentele subjecten waren vier van de acht groepsleden medewerkers van de onderzoeker, die unaniem een verkeerd antwoord gaven. Deze experimentele behandeling doet het percentage goede antwoorden dalen tot minder dan tweederde, wat suggereert dat mensen hun oordeel willen conformeren aan dat van anderen.

Gedurende het laatste decennium hebben goed ontworpen empirische studies ook het belang van conformisme op de werkvloer aangetoond. Ichino en Maggi (2000) laten zien dat werknemers zich minder ziekmelden en misdragen als hun collega's dit minder doen (zie ook Bradley et al. 2007). Falk en Ichino (2006) rapporteren dat, wanneer werknemers in koppels werken in plaats van individueel, de standaarddeviatie van hun output lager is en het niveau van hun output hoger. Bovendien

zijn het vooral de laag presterende werknemers die hun gedrag aanpassen als ze in een koppel werken. Mas en Moretti (2009) vinden dat kassamedewerkers hun prestaties verbeteren zodra een goed presterende kassamedewerker gaat meedraaien in hun shift. Net als in de studie van Falk en Ichino (2006) is dit effect veel sterker voor relatief slecht presterende medewerkers. De auteurs vinden daarnaast dat kassamedewerkers alleen hun prestaties verbeteren als dit geobserveerd kan worden door de nieuwe kassamedewerker. Op basis hiervan interpreteren de auteurs het gedrag van de kassamedewerkers in termen van sociale druk. Als laatste noem ik de studie van Bradler et al. (2013), die in een gecontroleerde werkomgeving laten zien dat het geven van publieke erkenning aan medewerkers de prestaties van medewerkers verhoogt. Zoals in Falk en Ichino (2006) en Mas en Moretti (2009) is dit effect het meest geprononceerd voor medewerkers die geen erkenning kregen, en dus leerden dat zij degenen waren die het slechtst gepresteerd hadden. Om deze reden interpreteren Bradler et al. (2013) hun resultaten deels in termen van de wens om te conformeren aan groepsnormen.

Tenslotte, hoewel de veronderstellingen van altruïsme en conformisme afwijken van de klassiek economische theorie, waarin verondersteld wordt dat mensen enkel hun geldelijk inkomen maximaliseren, handhaaf ik de veronderstelling dat mensen zich rationeel gedragen.[2] De theoretische modellen die hier ontwikkeld worden beschrijven altruïstische en conformistische voorkeuren daarom met behulp van nutsfuncties die gemaximaliseerd worden door de betrokken spelers. Bijgevolg lijken de modellen die in dit proefschrift ontwikkeld worden sterk op klassiek economische modellen, met dit verschil dat ze toelaten dat niet-financiële doelen worden nagestreefd.

Overzicht van het proefschrift

De hoofdstukken 2 en 3 van dit proefschrift hebben twee gemene delers. Ten eerste bestuderen beide hoofdstukken de implicaties van altruïsme tussen een werkne-

[2] Andreoni en Miller (2002) testen experimenteel de rationaliteit van altruïstische preferenties door de prijzen waartegen geld kan worden weggegeven aan een ander te variëren. De auteurs concluderen dat de meeste van hun subjecten keuzen maken die niet in strijd zijn met het Generalized Axiom of Revealed Preference.

mer en zijn werkgever of manager voor het optimaal gebruik van financiële prikkels. Ten tweede bestuderen beide hoofdstukken situaties waarin de prestaties van werknemers subjectief beoordeeld worden. In zulke gevallen kan de werkgever geen expliciet contract schrijven op basis van de prestaties van de werknemer, aangezien een subjectieve beoordeling niet geverifieerd kan worden door een rechter. In dit proefschrift worden twee alternatieve methoden om de werknemer te prikkelen bestudeerd. In het vervolg van dit hoofdstuk refereer ik naar een werkgever als principaal, naar een manager als supervisor, en naar een werknemer als agent.

In hoofdstuk 2 bestuderen Robert Dur en ik hoe relationele prestatie contracten beïnvloedt worden door altruïsme en afgunst tussen een principaal en agent. In een relationeel prestatie contract beoordeelt de principaal de prestaties van de agent subjectief, en belooft goede prestaties te belonen en/of slechte prestaties te straffen. Omdat een relationeel contract gebaseerd is op een subjectieve beoordeling door de principaal kan het niet door een rechter worden afgedwongen. Om de agent effectief te kunnen prikkelen moet een relationeel contract daarom geloofwaardig zijn. Dat wil zeggen dat de principaal bereidt moet zijn om beloften te honoreren en bedreigingen inderdaad uit te voeren. In het model maken we twee standaard veronderstellingen die geloofwaardige beloften en bedreigingen mogelijk maken: het spel wordt oneindig vaak herhaald, en de agent speelt een trigger strategy die voorschrijft om nooit meer hard te werken als de principaal het relationele contract heeft gebroken. Merk op dat de voorwaarde van geloofwaardigheid de beloften en bedreigingen van de principaal aanzienlijk kan beperken, en daarmee de motiverende kracht van een relationeel contract. We veronderstellen verder dat beide spelers een zekere mate van onderhandelingsmacht hebben, wat impliceert dat de totale waarde van de arbeidsrelatie tussen beide spelers verdeeld wordt.

In het model staan we de principaal toe twee manieren te gebruiken om de agent te prikkelen: een belofte van een bonus bij goede prestaties en een ontslagdreiging bij slechte prestaties. Ons eerste hoofdresultaat is dat zowel altruïsme van de principaal als van de agent de belofte van een bonus geloofwaardiger maken. De reden is dat het voor een altruïstische principaal minder kostbaar is om de bonus uit te betalen. Daarnaast maakt altruïsme tussen de spelers de arbeidsrelatie waardevoller.

Bijgevolg is het voor de principaal minder aantrekkelijk om een bonus onterecht in te houden, omdat dit de arbeidsrelatie op het spel zet. Het tweede hoofdresultaat is dat zowel altruïsme van de principaal als van de agent de geloofwaardigheid van een ontslagdreiging ondermijnen. De reden is dat, wanneer de spelers voldoende altruïstisch zijn, de arbeidsrelatie waardevol is, zelfs wanneer de agent er de kantjes vanaf loopt. Dreigen met ontslag is in dat geval niet geloofwaardig.

De hoofdresultaten leiden tot twee tegenintuïtieve comparatieve statica. Ten eerste, hoewel een altruïstische agent lagere financiële prikkels nodig heeft om hard te willen werken, kan sterker altruïsme toch leiden tot hogere bonussen. De reden is dat altruïsme vereist kan zijn om bonussen überhaupt geloofwaardig te maken. Dit resultaat is in tegenstelling tot de bestaande theoretische literatuur, die voorspelt dat altruïsme en financiële prikkels substituten zijn (zie bijv. Casadesus-Masanell 2004, Sliwka 2007, Shchetinin 2010, en Non 2012). Ten tweede, hoewel altruïsme impliceert dat spelers gedeeltelijk van elkaars welzijn genieten, kan het toch zo zijn dat een marginale stijging in altruïsme beide spelers slechter af maakt. Dit kan het geval zijn wanneer een stijging in altruïsme een ontslagdreiging ongeloofwaardig maakt. Het wegvallen van de ontslagdreiging als geloofwaardige prikkel kan impliceren dat de agent minder hard zal werken, waardoor de productiewaarde van de arbeidsrelatie daalt.

Terwijl hoofdstuk 2 altruïsme tussen een principaal en agent bestudeert, is hoofdstuk 3 gericht op de situatie waar een supervisor altruïstisch is ten opzichte van een agent. In het model dat ik ontwikkel is het de taak van de supervisor om de prestaties van de agent te beoordelen en dit te rapporteren aan de principaal. De principaal prikkelt de agent door een bonus te koppelen aan een hoge prestatiebeoordeling van de supervisor. De supervisor is een van twee types: egoïstisch, of altruïstisch naar de agent. De agent observeert het type van de supervisor, maar de principaal niet. Net zoals in hoofdstuk 2, veronderstel ik dat het spel oneindig herhaald wordt, en dat de agent een trigger strategie speelt die voorschrijft om nooit meer hard te werken als de supervisor een verkeerde prestatiebeoordeling heeft gegeven. Ik veronderstel verder dat er een grote pool van supervisors en agenten is. Iedere periode kan de principaal aan één supervisor en één agent een contract aanbieden. Tenslotte, het

contract voor de supervisor is uniform, d.w.z., de principaal kan de supervisor geen menu van contracten aanbieden.

In het model zoals hierboven beschreven kan het altruïsme van de supervisor de motivatie van de agent beïnvloeden. De reden is dat een altruïstische supervisor om het welzijn van de agent geeft, en daarom een prikkel heeft om hoge beoordelingen te geven ongeacht de werkelijke prestatie van de agent. Echter, als de agent dit verwacht heeft hij geen reden om hard te werken. De principaal moet de supervisor daarom prikkelen om correcte prestatiebeoordelingen te geven. In hoofdstuk 3 bestudeer ik hoe de principaal dit kan bereiken door optimale prestatie contracten aan de supervisor en de agent aan te bieden. De twee hoofdresultaten zijn als volgt.

Ten eerste, in tegenstelling tot de bestaande literatuur (bijv. Prendergast en Topel 1996 en Giebe en Gürtler 2012), laat ik zien dat het altruïsme van een supervisor een netto prikkel kan vormen om prestaties correct te beoordelen, in plaats van te hoog. Zoals in de hierboven genoemde modellen is het zo dat altruïsme een prikkel geeft om prestaties te hoog te beoordelen. Echter, in tegenstelling tot deze modellen laat ik zien dat altruïsme ook een prikkel geeft om prestaties correct te beoordelen. De reden is dat, vanwege altruïsme, de supervisor een surplus geniet van zijn baan. Dit surplus kan niet aan de supervisor worden onttrokken door zijn loon te verlagen, omdat dit betekent dat het egoïstische type niet langer de baan wil accepteren. Vervolgens laat ik zien dat het optimale contract impliceert dat de supervisor niet langer een surplus geniet nadat hij een incorrecte prestatiebeoordeling geeft. Het te hoog beoordelen van prestaties leidt in de toekomst dus tot verlies van het surplus dat bestaat vanwege de altruïstische relatie met de agent. Als de supervisor voldoende om de toekomst geeft, is dit laatste effect sterker dan het eerste, zodat altruïsme een netto prikkel geeft om prestaties correct te beoordelen. Ten tweede, het kan optimaal zijn om het contract van de supervisor zodanig op te stellen dat het slechts voor één van de typen aantrekkelijk is. De cruciale eigenschap van een dergelijke screening contract is dat het type van de supervisor onthuld wordt zodra het contract wordt aangenomen. Ik laat zien dat het bezitten van deze informatie de principaal in staat stelt om de supervisor goedkoper te prikkelen om correct prestaties te beoordelen. Als gevolg kan het aantrekkelijk zijn om het screening contract aan te bieden, zelfs

als dit de kans verkleint om de vacature voor supervisor te vervullen.

Hoofdstuk 4 van dit proefschrift bestudeert een andere motivatie op de werkvloer, namelijk de voorkeur om gedrag te conformeren aan sociale normen. Gegeven dat agenten hun gedrag conformeren aan sociale normen, ontstaat de vraag hoe organisaties winstmaximaliserende normen ontwikkelen. De analyse in hoofdstuk 4 biedt antwoorden op deze vraag. De innovatie van de studie is de veronderstelling dat normen binnen organisaties beïnvloedt worden door het gedrag van superieuren. Het doel van de studie is om de implicaties van deze assumptie voor norm creatie in hiërarchische organisaties bloot te leggen. Ik ontwikkel daarom een drielaags agency model (principaal-supervisor-agent), waarin elke speler zich inzet om waarde te genereren voor de onderneming van de principaal. De supervisor en de agent ondervinden psychologische kosten als hun inzet lager is dan de norm voor hun inzet. De inzet van de principaal vormt de norm voor de supervisor, wiens inzet op haar beurt de norm vormt voor de agent.

Het hoofdresultaat van hoofdstuk 4 is dat in hiërarchische ondernemingen 'norm erosie' optreedt. Met norm erosie bedoel ik dat normen lager zijn in lagere rangen van de organisatie. Het mechanisme dat norm erosie tot stand brengt is eenvoudig. Wanneer de supervisor zijn inzet bepaalt houdt hij rekening met het voorbeeld van de principaal, maar voldoet daar nooit volledig aan. De reden is dat het individueel optimaal is om de psychologische kosten van ondermaatse prestaties af te ruilen tegen de baten van minder hard werken. Als gevolg geeft de supervisor een lager voorbeeld aan de agent dan het voorbeeld dat hij zelf kreeg.

De implicatie van norm erosie is dat de norm voor de agent relatief laag is. Ik laat zien dat de principaal optimaal corrigeert voor norm erosie door zelf een hoger voorbeeld te geven. Echter, deze strategie leidt tot twee soorten kosten. Ten eerste kost het veel moeite voor de principaal zelf om de lat hoog te leggen. Ten tweede wordt de supervisor met een hoge norm geconfronteerd. Deze hoge norm leidt in evenwicht tot hoge kosten van inzet en hoge psychologische kosten van ondermaatse prestaties. Voor deze kosten moet de supervisor gecompenseerd worden in de vorm van een hoger loon. Norm erosie leidt dus tot drie inefficiënties in norm creatie: de inzet van de principaal is inefficiënt hoog, de norm voor de supervisor is inefficiënt

hoog, en de norm voor de agent is inefficiënt laag.

Het laatste deel van hoofdstuk 4 bestudeert hoe hiërarchische organisaties de inefficiënties in norm creatie kunnen minimaliseren. Een heldere implicatie van de analyse is dat het optimaal is om de mate van hiërarchie te beperken, aangezien dit de mogelijkheden voor norm erosie vermindert. Ten tweede is het optimaal om medewerkers met een hoge gevoeligheid voor sociale normen te promoveren. Dergelijke medewerkers conformeren hun gedrag sterker aan het voorbeeld van de principaal, wat de mate van norm erosie vermindert. Tenslotte laat ik zien dat het optimaal is om de span of control van supervisors te verstoren. De intuïtie is als volgt. In het model veronderstel ik dat er een exogeen gegeven supervisie technologie is die de optimale verhouding van supervisors tot agenten beïnvloedt. Echter, wanneer de principaal bepaalt hoeveel supervisors en agenten in te huren houdt zij ook rekening met de sociale normen die in evenwicht zullen ontstaan. Zoals hierboven uitgelegd is de norm voor supervisors altijd inefficiënt hoog, terwijl de norm voor agenten inefficiënt laag is. De mate van deze inefficiënties kan verschillen voor supervisors en agenten. Sterker, de principaal zorgt er in het optimum voor dat de inefficiëntie relatief klein is voor het type medewerker dat relatief veel wordt ingehuurd. De principaal bereikt dit door haar inzet te verlagen in het aantal supervisors en te verhogen in het aantal agenten. Hieruit volgt dat, als de supervisie technologie een hoge (lage) supervisor span of control aantrekkelijk maakt, agenten (supervisors) een relatief efficiënte norm hebben, waardoor agenten (supervisors) relatief aantrekkelijk zijn om in te huren. Als gevolg verstoren organisaties optimaal de span of control van supervisors.

Bibliography

Aggarwal, R.K. and A.A. Samwick (2003), "Performance Incentives Within Firms: The Effect of Managerial Responsibility." *Journal of Finance*, 58(4): 1613-1649.

Akerlof, G.A. (1980), "A Theory of Social Custom, of Which Unemployment May be One Consequence." *Quarterly Journal of Economics*, 94(4): 749-775.

Akerlof, G.A. (1982), "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics*, 97(4): 543-569.

Akerlof, G.A. and R.E. Kranton (2005), "Identity and the Economics of Organizations." *Journal of Economic Perspectives*, 19(1): 9-32.

Andreoni, J. and J. Miller (2002), "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70(2): 737-753.

Andreoni, J., W.T. Harbaugh, and L. Vesterlund (2008), "Altruism in Experiments." In: S.N. Durlauf and L.E. Blume (Eds.), *The New Palgrave Dictionary of Economics Online*, Palgrave Macmillan.

Arce, D.G. (2010), "Agency and Social Preferences." Mimeo, University of Texas at Dallas.

Asch, S.E. (1951), "Effects of Group Pressure Upon the Modification and Distortion of Judgment." In: Guetzkow, H. (Ed.), *Groups, Leadership and Men*, Pittsburg: Carnegie Press.

Baker, G., R. Gibbons, and K.J. Murphy (1994), "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics*, 109(4): 1125-1156.

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

Bartling, B. (2011), "Relative Performance or Team Evaluation? Optimal Contracts for Other-Regarding Agents." *Journal of Economic Behavior and Organization*, 79(3): 183-193.

Bartling, B. and F. Von Siemens (2010a), "The Intensity of Incentives in Firms and Markets: Moral Hazard with Envious Agents." *Labour Economics*, 17(2): 598-607.

Bartling, B. and F. Von Siemens (2010b), "Equal Sharing Rules in Partnerships." *Journal of Institutional and Theoretical Economics*, 166(2): 299–320.

Berger, J., C. Herbertz, and D. Sliwka (2011), "Managerial Incentives and Favoritism in Promotion Decisions: Theory and Field Evidence." IZA Discussion Paper, No. 5543.

Bergstrom, T.C. (1999), "Systems of Benevolent Utility Functions." *Journal of Public Economic Theory*, 1(1): 71-100.

Berman, E.M., J.P. West, and M.N. Richter (2002), "Workplace Relations: Friendship Patterns and Consequences (According to Managers)." *Public Administration Review*, 62(2): 217-230.

Bernheim, D.B. (1994), "A Theory of Conformity." *Journal of Political Economy*, 102(5): 841-877.

Besley, T. and M. Ghatak (2005), "Competition and Incentives with Motivated Agents." *American Economic Review*, 95(3): 616-636.

Bol, J.C. (2011), "The Determinants and Performance Effects of Managers' Performance Evaluation Biases." *Accounting Review*, 85(5): 1549-1575.

Bolton, G.E. and A. Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166-193.

Borzaga C. and S. Depedri (2005), "Interpersonal Relations and Job Satisfaction: Some Empirical Results in Social and Community Care Services." In: Gui B. and R. Sugden (Eds.), *Economics and Social Interaction: Accounting for Interpersonal Relations*, Cambridge: Cambridge University Press.

Bradler, C., R. Dur, S. Neckermann, and A. Non (2013), "Employee Recognition and Performance: A Field Experiment." Tinbergen Institute Discussion Paper, TI 2013-038/VII.

Bradley, S., C. Green, and G. Leeves (2007), "Worker Absence and Shirking: Evidence from Matched Teacher-School Data." *Labour Economics*, 14(3): 319-334.

Brown, M., A. Falk, and E. Fehr (2004), "Relational Contracts and the Nature of Market Interactions." *Econometrica*, 72(3): 747-780.

Brown, M., A. Falk, and E. Fehr (2012), "Competition and Relational Contracts: The Role of Unemployment as a Disciplinary Device." *Journal of the European Economic Association*, 10(4): 887-907.

Bull, C. (1987), "The Existence of Self-Enforcing Implicit Contracts." *Quarterly Journal of Economics*, 102(1): 147-159.

Campbell, C.M. and K.S. Kamlani (1997), "The Reasons for Wage Rigidity: Evidence From a Survey of Firms." *Quarterly Journal of Economics*, 112(3): 759-789.

Carlin, B.I. and S. Gervais (2009), "Work Ethic, Employment Contracts, and Firm Value." *Journal of Finance*, 64(2): 758-821.

Carmichael, L. (1983), "The Agent-Agents Problem: Payment by Relative Output." *Journal of Labor Economics* 1(1): 50-65.

Casadesus-Masanell, R. (2004), "Trust in Agency." *Journal of Economics and Management Strategy*, 13(3): 375-404.

Cordero Salas, P. (2011), "The Role of Bargaining Power in Relational Contracts." SSRN Working Paper.

Cordero Salas, P. and B. Roe (2012), "The Role of Cooperation and Reciprocity in Structuring Carbon Sequestration Contracts in Developing Countries." *American Journal of Agricultural Economics*, 94(2): 411-418.

Delfgaauw, J. and R. Dur (2007), "Signaling and Screening of Worker's Motivation." *Journal of Economic Behavior and Organization*, 62(4): 605–624.

Delfgaauw, J. and R. Dur (2008), "Incentives and Workers' Motivation in the Public Sector." *Economic Journal*, 118(525): 171–191.

DellaVigna, S., J.A. List, and U. Malmendier (2012), "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics*, 127(1): 1-56.

Dur, R. (2009), "Gift Exchange in the Workplace: Money or Attention?" *Journal of the European Economic Association*, 7(2-3), 550-560.

Dur, R. and A. Glazer (2008), "Optimal Contracts When a Worker Envies His Boss." *Journal of Law, Economics, and Organization*, 24(1): 120-137.

Dur, R. and J. Sol (2010), "Social Interaction, Co-Worker Altruism, and Incentives." *Games and Economic Behavior*, 69(2): 293-301.

Ederer, F. and A. Patacconi (2010), "Interpersonal Comparison, Status and Ambition in Organizations." *Journal of Economic Behavior and Organization*, 75(2): 348-363.

Englmaier, F. and S. Leider (2012), "Contractual and Organizational Structure with Reciprocal Agents." *American Economic Journal: Microeconomics*, 4(2): 146–183.

Englmaier, F. and A. Wambach (2010), "Optimal Incentive Contracts under Inequity Aversion." *Games and Economic Behavior*, 69(2): 312-328.

Falk, A. and A. Ichino (2006), "Clean Evidence on Peer Effects." *Journal of Labor Economics*, 24(1): 39-57.

Fehr, E., A. Klein, and K. Schmidt (2007), "Fairness and Contractual Design." *Econometrica* 75(1): 121-154.

Fehr, E. and K. Schmidt (1999), "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.

Fehr, E. and K. Schmidt (2000), "Fairness, Incentives, and Contractual Choices." *European Economic Review*, 44(4-6): 1057–1068.

Fischer, P. and S. Huddart (2008), "Optimal Contracting with Endogenous Social Norms." *American Economic Review*, 98(4): 1459-1475.

Gibbons, R. (1998), "Incentives in Organizations." *Journal of Economic Perspectives*, 12(4): 115-132.

Gibbs, M. (2012), "Design and Implementation of Pay for Performance." IZA Discussion Paper No. 6322.

Gibbs, M., K.A. Merchant, W.A. Van der Stede, and M.E. Vargus (2004), "Determinants and Effects of Subjectivity in Incentives." *Accounting Review*, 79(2): 409-436.

Giebe, T. and O. Gürtler (2012), "Optimal Contracts for Lenient Supervisors." *Journal of Economic Behavior and Organization*, 81(2): 403-420.

Golman, R. and S. Bhatia (2012), "Performance Evaluation Inflation and Compression." *Accounting, Organizations and Society*, 37(8): 534-543.

Grund, C. and D. Sliwka (2005), "Envy and Compassion in Tournaments." *Journal of Economics and Management Strategy*, 14(1): 187-207.

Grund, C. and J. Przemeck (2012), "Subjective Performance Evaluation and Inequality Aversion." *Applied Economics*, 44(17): 2149-2155.

Heinle, M.S., C. Hofmann, and A.H. Kunz (2012), "Identity, Incentives, and the Value of Information." *Accounting Review*, 87(4): 1309-1334.

Hermalin, B.E. (1998), "An Economic Theory of Leadership." *American Economic Review*, 88(5): 1188-1206.

Huck, S., D. Kübler, and J. Weibull (2012), "Social Norms and Economic Incentives in Firms." *Journal of Economic Behavior and Organization*, 83(2): 173-185.

Ichino, A. and G. Maggi (2000), "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm." *Quarterly Journal of Economics*, 115(3): 1057-1090.

Itoh, H. (2004), "Moral Hazard and Other-Regarding Preferences." *Japanese Economic Review*, 55(1): 18-45.

Jawahar, I.M. and C.R. Williams (1997), "Where All The Children Are Above Average: The Performance Appraisal Purpose Effect." *Personnel Psychology*, 50(4): 905-926.

Kahneman, D., A.B. Krueger, D.A. Schkade, N. Schwarz, and A.A. Stone (2004), "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method." *Science*, 306(5702): 1776-1780.

Kamphorst, J.J.A. and O.H. Swank (2012), "The Role of Performance Appraisals in Motivating Employees." Tinbergen Institute Discussion Paper, TI2012-034/1.

Kandel, E. and E.P. Lazear (1992), "Peer Pressure and Partnerships." *Journal of Political Economy*, 100(4): 801-817.

Kosfeld, M. and F.A. Von Siemens (2009), "Worker Self-selection and the Profits from Cooperation." *Journal of the European Economic Association*, 7(2-3): 573–582.

Kosfeld, M. and F.A. Von Siemens (2011), "Competition, Cooperation, and Corporate Culture." *RAND Journal of Economics*, 42(1): 23–43.

Kragl, J. (2009), "Group vs. Individual Performance Pay in Relational Employment Contracts when Workers Are Envious." European Business School Research Paper No. 09-09.

Kragl, J. (2011), "Individual vs. Relative Performance Pay with Envious Workers and Non-verifiable Performance." European Business School Research Paper No. 11-04.

Kragl, J. and J. Schmid (2009), "The Impact of Envy on Relational Incentive Contracts." *Journal of Economic Behavior and Organization*, 72(2): 766-779.

Landry, C.E., A. Lange, J.A. List, M.K. Price, and N.G. Rupp (2006), "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment." *Quarterly Journal of Economics*, 121(2): 747-782.

Laffont, J.-J. and D. Martimort (2002), *The Theory of Incentives*, Princeton and Oxford: Princeton University Press.

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

Lazear, E.P. and P. Oyer (2013), "Personnel Economics." In: Gibbons, R. and J. Roberts (Eds.), *Handbook of Organizational Economics*, Princeton and Oxford: Princeton University Press.

Ledyard, J.O. (1995), "Public Goods: A Survey of Experimental Research." In: Kagel, J. and A. Roth (Eds.), *Handbook of Experimental Economics*, New Jersey: Princeton University Press.

Lee, S. and P. Persson (2011), "Authority versus Loyalty: Social Incentives and Governance." NYU Working Paper, No. FIN-10-001.

Levin, J. (2003), "Relational Incentive Contracts." *American Economic Review*, 97(3): 835-857.

Levitt, S. and J.A. List (2007), "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21(2): 153-174.

MacLeod, W.B. and J.M. Malcomson (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica*, 57(2): 447-480.

MacLeod, W.B. and J.M. Malcomson (1998), "Motivation and Markets." *American Economic Review*, 88(3): 388-411.

MacLeod, W.B. and D. Parent, (2000) "Job Characteristics and the Form of Compensation." In: Polachek, S.W. and J. Robst (Eds.), *Research in Labor Economics,* Vol. 18, London: JAI Press.

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

Medoff, J.L. and K. G. Abraham (1980), "Experience, Performance, and Earnings." *Quarterly Journal of Economics*, 95(4): 703-736.

Millward, N., M. Stevens, D. Smart, and W.R. Hawes (1992), *Workplace Industrial Relations in Transition: The ED/ESRC/PSI/ACAS Surveys*, Aldershot: Dartmouth.

Moerbeek, H.H.S. and A. Need (2003), "Enemies at Work: Can They Hinder Your Career?" *Social Networks*, 25(1): 67-82.

Moers, F. (2005), "Discretion and Bias in Performance Evaluation: The Impact of Diversity and Subjectivity." *Accounting, Organizations and Society*, 30(1): 67-80.

Murphy, P.E. and G. Enderle (1995), "Managerial Ethical Leadership: Examples do Matter." *Business Ethics Quarterly*, 5(1): 117-128.

Müller, D. and P. Weinschenk (2011), "Rater Bias in Performance Appraisal: on Horns, Halos, and Incentive Provision." Mimeo, University of Bonn.

Neilson, W.S. and J. Stowe (2010). "Piece-Rate Contracts for Other-Regarding Workers." *Economic Inquiry*, 48(3): 575-586.

Non, A. (2012), "Gift-Exchange, Incentives, and Heterogeneous Workers." *Games and Economic Behavior*, 75(1): 319-336.

Posner, B.Z. and W.H. Schmidt (1992), "Values and the American Manager: An Update Updated." *California Management Review*, 34(3): 80-94.

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

Prendergast, C. and R.H. Topel (1996), "Favoritism in Organizations." *Journal of Political Economy*, 104(5): 958-978.

Raff, D.M.G. and L.H. Summers (1987), "Did Henry Ford Pay Efficiency Wages?" *Journal of Labor Economics*, 5(4): S57-S86.

Rey-Biel, P. (2008). "Inequity Aversion and Team Incentives." *Scandinavian Journal of Economics*, 108(2), 297-320.

Rhoades, L. and R. Eisenberger (2002), "Perceived Organizational Support: A Review of the Literature." *Journal of Applied Psychology*, 87(4): 698-714.

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

Shapiro, C. and J.E. Stiglitz (1984), "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, 74(3): 433-444.

Shchetinin, O. (2010), "Altruism and Career Concerns." Gothenborg University, Working Papers in Economics 427.

Shearer, B. (2004), "Piece Rates, Fixed Wages and Incentives: Evidence From a Field Experiment." *Review of Economics Studies*, 71(2): 513-534.

Sliwka, D. (2007), "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *American Economic Review*, 97(3): 999-1012.

Sol, J. (2010), "Peer Evaluation: Incentives and Co-Worker Relations." Tinbergen Institute Discussion Paper, TI2010-055/1.

Spence, M. (1973), "Job Market Signalling." *Quarterly Journal of Economics*, 87(3): 355-374.

Stevens, D.E. and A. Thevaranjan (2010), "A Moral Solution to the Moral Hazard Problem." *Accounting, Organizations and Society*, 35(1): 125-139.

Thiele, V. (2013), "Subjective Performance Evaluation, Collusion, and Organizational Design." *Journal of Law, Economics, and Organization*, 29(1): 35-59.

Treviño, L.K., G.R. Weaver, D.G. Gibson, and B.L. Toffler (1999), "Managing Ethics and Legal Compliance: What Works and What Hurts." *California Management Review*, 41(2): 131-151.

Treviño, L.K. and G.R. Weaver (2003), "Ethics and the Broader Organizational Context: Ethical Climate and Ethical Culture." In: Treviño, L.K. and G.R. Weaver (Eds.), *Managing Ethics in Business Organizations*, Stanford, CA: Stanford University Press.

Von Siemens, F.A. (2011), "Heterogeneous Social Preferences, Screening and Employment Contracts." *Oxford Economic Papers*, 63(3): 499-522.

Von Siemens, F.A. (2012), "Social Preferences, Sorting and Competition." *Scandinavian Journal of Economics*, 114(3): 780-807.

Wayne, S.J., L.M. Shore, and R.C. Liden (1997), "Perceived Organizational Support and Leader–Member Exchange: A Social Exchange Perspective." *Academy of Management Journal*, 40(1): 82–111.

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:

530 R. LEGERSTEE, Evaluating Econometric Models and Expert Intuition
531 M.R.C. BERSEM, Essays on the Political Economy of Finance
532 T. WILLEMS, Essays on Optimal Experimentation
533 Z. GAO, Essays on Empirical Likelihood in Economics
534 J. SWART, Natural Resources and the Environment: Implications for Economic Development and International Relations
535 A. KOTHIYAL, Subjective Probability and Ambiguity
536 B. VOOGT, Essays on Consumer Search and Dynamic Committees
537 T. DE HAAN, Strategic Communication: Theory and Experiment
538 T. BUSER, Essays in Behavioural Economics
539 J.A. ROSERO MONCAYO, On the Importance of Families and Public Policies for Child Development Outcomes
540 E. ERDOGAN CIFTCI, Health Perceptions and Labor Force Participation of Older Workers
541 T.WANG, Essays on Empirical Market Microstructure
542 T. BAO, Experiments on Heterogeneous Expectations and Switching Behavior
543 S.D. LANSDORP, On Risks and Opportunities in Financial Markets
544 N. MOES, Cooperative Decision Making in River Water Allocation Problems
545 P. STAKENAS, Fractional Integration and Cointegration in Financial Time Series
546 M. SCHARTH, Essays on Monte Carlo Methods for State Space Models
547 J. ZENHORST, Macroeconomic Perspectives on the Equity Premium Puzzle
548 B. PELLOUX, the Role of Emotions and Social Ties in Public On Good Games: Behavioral and Neuroeconomic Studies
549 N. YANG, Markov-Perfect Industry Dynamics: Theory, Computation, and Applications
550 R.R. VAN VELDHUIZEN, Essays in Experimental Economics
551 X. ZHANG, Modeling Time Variation in Systemic Risk
552 H.R.A. KOSTER, The Internal Structure of Cities: the Economics of Agglomeration, Amenities and Accessibility
553 S.P.T. GROOT, Agglomeration, Globalization and Regional Labor Markets: Micro Evidence for the Netherlands
554 J.L. MÖHLMANN, Globalization and Productivity Micro-Evidence on Heterogeneous Firms, Workers and Products
555 S.M. HOOGENDOORN, Diversity and Team Performance: A Series of Field Experiments

556 C.L. BEHRENS, Product Differentiation in Aviation Passenger Markets: The Impact of Demand Heterogeneity on Competition
557 G. SMRKOLJ, Dynamic Models of Research and Development
558 S. PEER, The Economics of Trip Scheduling, Travel Time Variability and Traffic Information
559 V. SPINU, Nonadditive Beliefs: From Measurement to Extensions
560 S.P. KASTORYANO, Essays in Applied Dynamic Microeconometrics
561 M. VAN DUIJN, Location, Choice, Cultural Heritage and House Prices
562 T. SALIMANS, Essays in Likelihood-Based Computational Econometrics
563 P. SUN, Tail Risk of Equidity Returns
564 C.G.J. KARSTEN, The Law and Finance of M&A Contracts
565 C. OZGEN, Impacts of Immigration and Cultural Diversity on Innovation and Economic Growth
566 R.S. SCHOLTE, The Interplay Between Early-Life Conditions, Major Events and Health Later in Life
567 B.N.KRAMER, Why Don't They Take a Card? Essays on the Demand for Micro Health Insurance
568 M. KILIÇ, Fundamental Insights in Power Futures Prices
569 A.G.B. DE VRIES, Venture Capital: Relations with the Economy and Intellectual Property
570 E.M.F. VAN DEN BROEK, Keeping up Appearances
571 K.T. MOORE, A Tale of Risk: Essays on Financial Extremes
572 F.T. ZOUTMAN, A Symphony of Redistributive Instruments
573 M.J. GERRITSE, Policy Competition and the Spatial Economy
574 A. OPSCHOOR, Understanding Financial Market Volatility
575 R.R. VAN LOON, Tourism and the Economic Valuation of Cultural Heritage
576 I.L. LYUBIMOV, Essays on Political Economy and Economic Development
577 A.A.F. GERRITSEN, Essays in Optimal Government Policy
578 M.L. SCHOLTUS, The Impact of High-Frequency Trading on Financial Markets
579 E. RAVIV, Forecasting Financial and Macroeconomic Variables: Shrinkage, Dimension reduction, and Aggregation

Performance pay can motivate employees, but money is not the only motivation in the workplace. Altruism, which means that someone enjoys the well-being of someone else, can also provide a powerful motivation. The first part of this thesis studies theoretically how altruism between an employee and his superior affects the optimal use of monetary incentives. Among others, the analysis reveals how altruism influences the credibility of monetary incentive schemes, and how altruistic managers can be prevented from being lenient in evaluating employee performance. The second part of this thesis focuses on motivation stemming from the desire to conform to social norms. I develop a theoretical model in which it is assumed that norms for employees are affected by the example of superiors. The analysis sheds light on how superiors take the norm-setting aspect of their behavior into account, and derives implications for the optimal design of organizations.

Jan Tichem (1987) graduated in Economics at Erasmus University Rotterdam in 2010. Afterwards, he joined the Tinbergen Institute as a PhD-student at the Department of Economics at Erasmus University. Currently, he works at the Office of the Chief Economist at Authority for Consumers & Markets (ACM). A part of this thesis is forthcoming in Journal of Economics & Management Strategy.