

Incentives at work

ISBN: 978 90 3610 576 7

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

This book is no. **746** of the Tinbergen Institute Research Series, established through cooperation between Rozenberg publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Incentives at Work

Prikkels aan het werk

Proefschrift

ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

op gezag van de

rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens het besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 7 november 2019 om 13.30 uur

door

Maarten Niels Souverijn

geboren op zondag 7 juni 1987

te Leiden

Erasmus University Rotterdam



Promotiecommissie

Promotor:	Prof.dr. A.J. Dur
Overige leden:	Prof.dr. A. Schöttner Dr. J.T.R. Stoop Prof.dr. O.H. Swank
Co-promotor	Dr. J. Delfgaauw

Preface

No journey to a doctoral thesis is easy. Mine certainly was not. Fortunately I have been the benefactor of plenty of support along the way, for which I am most grateful. I would like to thank my colleagues, supervisors, friends and family for all their support.

Josse, I could not have hoped for a better supervisor. We have had plenty of good discussions and I have benefitted immensely from your advice and constructive criticism. Getting a draft scrabbled with notes was not always fun, but always for the better. Most importantly, your ability to relate economic theory to the real world is much appreciated.

Robert Dur, for acting as promotor and co-author, bringing a level of enthusiasm that is quite inspiring.

I must also thank the staff of the department. The discussions during and after seminars were most insightful. Special thanks go to Jurjen Kamphorst whose supervision of my bachelor thesis and courses kindled my love for economics greatly.

Of course it is common knowledge that most socialization takes place among peers. Academia is no different. The network of officemates and PhD colleagues provided much needed support. In particular my officemates Pengfei and Ruben with whom I enjoyed many conversations on teaching, research and the philosophy of economics. Oke, Jan, Robin and Harro cannot be forgotten for the many interesting conversations both on- and off-topic. I must especially thank Max,

my co-author and the primary victim of philosophizing in my final year at H8. To my friends Guido, Stefan, Maxim and Vincent: our weekly meetups were more important for this than you imagine.

Finally and most importantly I must thank my family. Mom, dad, I could never have accomplished this without your support.

This thesis has been long in the making. There are plenty of reasons, but no excuse. I realize that it has been hard to watch the struggle and must once more thank everyone profusely for their support. I could not have done it without you.

Contents

Preface	i
Introduction	1
1 Incentives and Limited Measurement	13
1.1 Introduction	13
1.2 Related literature	16
1.3 The model	20
1.4 First-best Benchmark	22
1.5 Analysis	23
1.5.1 Output based measured performance	23
1.5.2 Input based measured performance	25
1.6 Discussion and concluding remarks	32
1.A Appendix	35
2 Biased Supervision	45
2.1 Introduction	45
2.2 The model	52
2.3 Benchmarks	56
2.3.1 Complete information	56
2.3.2 Pure objective performance pay	56
2.3.3 Pure subjective performance pay	58

2.3.4	Pure subjective versus pure objective pay for performance	61
2.4	Subjective and objective performance pay combined	62
2.5	Generalists versus specialists as supervisors	76
2.6	Concluding remarks	78
2.A	Appendix	79
3	Team Incentives, Task Assignment, and Performance: A Field Experiment	93
3.1	Introduction	93
3.2	Experimental Context and Design	98
3.2.1	Experimental context	98
3.2.2	Experimental Design	100
3.2.3	Surveys	103
3.2.4	Assignment procedure	106
3.3	Estimation	107
3.4	Descriptive statistics	109
3.5	Results	117
3.6	Discussion	121
3.A	Appendix	125
4	Goal Setting and Raising the Bar	129
4.1	Introduction	129
4.2	Experimental Set-up and Data Description	133
4.2.1	Experimental Context	133
4.2.2	Experimental Design	135
4.2.3	Assignment Procedure	138
4.3	Predictions	139
4.4	Empirical Strategy	142
4.5	Descriptive Statistics	144
4.6	Results	152
4.6.1	Total effect	153

Contents

4.6.2	Intensive margin	154
4.6.3	Extensive margin	156
4.6.4	Further results	157
4.7	Conclusion	165
4.A	Mentor instructions	168
4.A.1	Goal treatment	168
4.A.2	Raise treatment	170
Summary		173
Samenvatting		179
Bibliography		185

List of Figures

2.1	Preferred pure incentive mechanism.	62
2.2	Feasible supervisor demands	66
2.3	Case where supervisor is more biased than performance measure. .	70
2.4	Case where measured performance is too unreliable.	73
3.1	Task assignment and efficiency, store averages	100
3.2	Example of intermediate ranking provided to stores	105
3.3	Average sales and average targeted sales	110
3.4	Performance of treatment and control stores.	111
4.1	Flowchart of treatments	137
4.2	Randomization procedure	140
4.3	Initial goals set by students	149

List of Tables

3.1	Survey questions on task assignment	104
3.2	Store characteristics, by first-period assignment	110
3.3	Results of first and second survey, store averages	112
3.4	Correlations between drivers of task allocation in first survey, store- level averages	114
3.5	Estimation performance effect (intermediate weeks discarded) . . .	118
3.6	Task allocation estimates, store level	120
3.7	Task allocation estimates, worker level	120
3.8	Job satisfaction	122
A.1	Characteristics of first and second survey (non-)respondents, by treatment	125
A.2	Store characteristics and survey response	126
A.3	Average worker responses by survey participation and treatment .	127
A.4	Store average responses by survey participation and treatment . .	128
4.1	Descriptives by assigned treatment	146
4.2	Students asked and not asked to set a goal within the treatment groups	147
4.3	Students asked and not asked to raise their goal in the raise treatment	148
4.4	Students who have a goal in mind when asked	150
4.5	Mentor estimates are accurate	150

4.6	Students who set and do not set goals when asked to set a goal . .	151
4.7	Students who accept and reject raise when asked to raise their goal	152
4.8	Previous course results in Dutch and international programmes . .	152
4.9	Intention to treat effect: total effect	155
4.10	Intention to treat effect: intensive margin	155
4.11	Intention to treat effect: dropout	157
4.12	Intention to treat effect: maximum grade including resit	157
4.13	ITT interaction treatment with GPA in preceding block	159
4.14	ITT interaction treatment with gender of mentor	160
4.15	ITT interaction treatment with experience of mentor	162

Introduction

The study of economics is the study of choice. All choice is influenced by the context in which it is made, the options, available information and the consequences of choices, known to economists as the incentive structure. A premise of economic theory is that by changing the incentives you can change behaviour. How to incentivize people such that they do as desired is one of the core questions in economics, life, and this thesis in particular. Since we all rely upon others for our success this question is regularly pondered by managers, teachers, coaches, parents, shareholders and consumers. We are all aware of the abstract theoretical solution: use incentives to align interests. However more practical examples tailored to existing incentive structures are needed. For instance, how should a worker be compensated? How does this change if he performs multiple tasks, if his output and/or inputs cannot be observed or if he works in a team?

Incentive mechanisms come in many shapes and sizes, running from objective and contractual to subjective, input- to output-based, for individuals or groups, direct to deferred, monetary to non-monetary and absolute to relative, in addition to possible combinations of these. Each incentive mechanism has its own advantages and disadvantages that make it more or less suitable in particular contexts. This makes that there is no one-size-fits-all solution. Indeed, the central theme of this thesis is that incentivizing others is possible but not straightforward. Studying incentives is important as examples abound of incentive schemes that backfire, lowering performance rather than raising it.¹

¹Examples include teachers teaching to the test if rewarded for student performance, account

Theory has traditionally been ahead of empirical evidence in this field of economics. Game theoretic analysis provides clear-cut results by forcing the specification of assumptions and taking them to their logical conclusion. Theory provides internal consistency but can lack external validity since it is the ultimate ivory tower form of pursuing knowledge. In recent years however there has been an increasing effort to test incentive theory in the field by means of field experiments. This is done by randomizing the incentive structure amongst participants such that the effects of different incentives can be studied directly.

This thesis contains contributions to economic theory as well as field experiments testing theory. Theoretically I study optimal incentive schemes when measured performance is imperfectly related to actual performance in two different ways. First, I consider the case in which measured performance is more coarse than actual performance. Second, I consider the case in which a worker is evaluated by both a biased manager and a biased performance measure. In the field experiments I study the effect of changes in the incentive structure in real world settings. First, I study how a team incentive affects team performance and the assignment of tasks within teams. Second, I study how students respond when they are asked to set goals and are encouraged to raise their goal.

The remainder of this introduction is organized as follows. In the next section the basic theory of incentives is discussed. An overview of the chapters concludes the introduction.

Incentive theory

The fundamental problem of incentive theory is how to utilize the available information to induce efficient behaviour from contracting parties when such behaviour cannot be contracted upon directly. The efficiency of such actions absent an optimized contract is in question because relationships tend to be underlain with conflict. For example, working hard benefits the employer but not necessarily managers focusing on sales not service, athletes using doping and workers shirking.

Introduction

essarily the employee. Stated simply the problem is one of aligning diverging interests and focussing the actions of all involved parties towards the overall goal of maximizing joint rents. Alignment can be brought about through explicit incentive contracts relying on hard (verifiable) information, or on implicit relational contracts using soft information such as subjective evaluation. Non-contractual means of aligning interests include the use of norms, goals, and reputation. If alignment of interests proves impossible exercising authority and control may be an option, for instance through a strict separation of tasks. The best solution depends on the specifics of the situation. This includes the information available, the stability of the environment, the tasks performed and whether it is an individual or team effort. The rest of this section discusses the primary lessons from this literature in more detail. Overviews of the literature are provided by Bolton and Dewatripont (2005) and Holmström (2017) among others.

The problem of utilizing the available information in order to induce efficient behaviour has been studied most intensely in single task principal-agent models. In these models a principal has an opportunity to earn economic rents but requires the efforts of an agent to capitalize on the opportunity. The efficient, first-best outcome is achieved if the agent's efforts maximize total surplus. In choosing efforts the agent is making a trade-off between the costs of his efforts and the rewards accruing to him as a result, as opposed to the effects of his efforts on all involved parties. This is the conflict of interests at the heart of the literature on incentives: private costs versus public benefits. The classic solution is to use incentive pay in order to induce the agent to deliver the desired amount of effort. This entails utilizing some sort of verifiable performance measure and tying compensation to measured performance. By paying the agent the full marginal product of effort his private benefit equals that of the public, solving the conflict of interest. In the language of incentive theory pay-for-performance contracts make the agent internalize the full benefits of his efforts such that a proper trade-off between the costs and benefits of effort can be made.²

²Note that pay-for-performance contracts can be construed in a variety of ways to ensure

The above demonstrates how pay-for-performance can solve the central problem in incentive theory. Using measured performance works well if it is directly related to actual performance, the actual efforts the agent puts in. However measured performance is often imperfect, in at least one of several ways. First, measured performance is often related not to input but to the more volatile output. This means that the agent is forced to accept the risk of an uncertain reward, undermining the incentive effect of pay-for-performance if the agent is risk-averse (eg. Shavell 1979). In addition the agent sometimes receives an unjust reward, lowering the willingness to offer rewards in the first place. The less direct the link between efforts and output and the more risk averse the agent the bigger this problem becomes.

Second, measured performance can neglect or overemphasize aspects of actual desired performance. This incentivizes the agent to under- and overdeliver these aspects of performance (eg. Holmström and Milgrom 1991, Baker 1992, 2002). One solution can be to add additional performance measures, up to the point where additional measures do not provide new information (eg. Holmström 1979, Feltham and Xie 1994). Still this may not yield the desired solution. For one, in some cases performance measurement may simply be impossible to get right. For instance in many cases agents work in teams. Even if team performance can be measured using it as a basis for pay does not incentivize team members on an individual level and hence does not get rid of free rider problems. Or consider dynamic environments where what constitutes proper behaviour can shift from one moment to the next.

An alternative to objective performance measurement is to subjectively evaluate performance. Subjective performance evaluation can be more accurate in terms of properly balancing the different aspects of worker performance (eg. Baker et al. 1994). In addition it offers more flexibility in terms of changing what is considered important during the contracted period. However subjective evaluation

that the agent only obtains the full marginal product on the margin. This ensures that the agent still makes the proper trade-off but does not capture all of the value of the opportunity, a necessity if others are to play a part in discovering and exploiting it as well.

Introduction

is not without drawbacks. First, it takes time and effort to monitor the agent. Hiring a supervisor to do the monitoring can result in the age old problem of monitoring the monitor (eg. Alchian & Demsetz 1972, Gibbs et al. 2004, Kamphorst and Swank 2012). Second, it can lack commitment and hence credibility from the agent's perspective as promised pay based on a good evaluation can be withheld at the last moment. Finally evaluation can be subject to favouritism (eg. Prendergast and Topel 1996, Bol 2011, Dur and Tichem 2015), extortion (eg. Laffont 1990, Vafai 2002, 2010), or collusion (eg. Tirole 1986, Thiele 2013). These factors can lower the efficacy of subjective performance pay, since they can result in an agent doing a good job not getting his just deserts, or in one receiving an unjust reward. Extortion by a supervisor can result in lower incentive strength or in the agent performing work that is not in the principal's interest. Some of these effects may be overcome by the concern for reputation. After all, if a promise of good evaluation is not kept an agent is unlikely to act based on a new promise, nor are others that are aware of the deception. As a result employment contracts can often be seen as a relational contract in which principal and agent work together for some time, but do not specify all contingencies (eg. Bull 1987, Levin 2003). Decision rights in case of eventualities govern the contract, but the decider in any one case needs to take the effects on the other into account.

In the discussion above we have considered ways in which agents can be incentivized to do the right things by using performance information. Thus we have considered the monitoring of outputs in order to obtain the right inputs. Instead one could imagine monitoring inputs directly. This is a great strategy for well defined jobs where it is precisely known what a worker needs to do. However it can be harder to define all the various little elements that a worker needs to do than to define what good performance is. In addition pay conditional on performance gives the agent some freedom in pursuing performance, allowing adaptation to changes in the environment and the exploration of new methods. It allows the agent to respond to the situation on the ground such that information is utilized directly in the production process. This advantage is greater the more uncer-

tainty there is ex-ante regarding what the agent should do, and the greater the information advantage of the agent. As a result basing rewards on performance can be a better choice than utilizing input based measures (eg. Prendergast 2002, Raith 2008).

Note that so far we have considered incentive pay as a mechanism to make agents do the right thing. This works by making agents feel the consequences of their actions by means of their remuneration. There are of course many other ways of making agents (feel) responsible for good behaviour. For one, the desire for a good reputation can be used. A reputation can be gained by acting properly or by achieving outcomes that are associated with either good behaviour or valued characteristics. A good reputation need not be valuable per se but can also be valuable because it opens up future opportunities. Regardless of whether the concern for reputation is an indirectly material concern or not, it does provide an incentive to pursue something other than immediate self-interest. Reputational concerns can provide incentives to work, to reveal information and can influence project choice. There is a rich literature on reputational concerns starting with the seminal contribution of Homström (1982).

Related to the concern for reputation is the notion of norm-following. People may follow norms, whether of society, an organization, or their direct colleagues. Norms are part of the culture of an organization, and can influence effort, task prioritization, and the handling of unexpected contingencies (eg. Akerlof 1980, Kandel and Lazear 1992, Mas and Moretti 2009). Norms and reputational concerns may help teams overcome the shirking problem where members put in less effort than is optimal for the team because the team shares the benefits rather than the benefits accruing to the individual. Another related mechanism that can make agents feel accountable for their results is goal setting. A goal can serve as a norm, whether publicly or privately, making attaining or at least not failing in reaching the goal attractive (eg. Suvorov and Van de Ven 2008, Koch and Nafziger 2011, and Hsiaw 2013). It should be noted however that one cannot set just any goal and expect it to be met, for a goal to be effective it must

Introduction

be attainable for the agent (eg. Locke and Latham 2002, 2006). The trick to entice the agent to high performance then is to set a goal that is attainable but somewhat difficult to reach.

Finally there is a whole range of factors influencing job performance that are related to how the task is being perceived. It can be very important for motivation that the task that has to be executed is being perceived as significant, that the task is worth doing. This can be because it plays an important role in the functioning of the organization, or because it affects others outside of the organization. The significance of good performance for others helps motivate by appealing to altruistic feelings of workers. The study of such factors are in its infancy in economics but the task significance literature is much more developed in psychology (eg. Hackman and Oldman 1976, Humphrey et al. 2007).

To summarize, incentive theory concerns itself with the problem of aligning interests. When interests are aligned all involved parties effectively work towards the common goal of joint utility maximization. When incentives are misaligned some interests necessarily receive more attention than required while others receive less, resulting in inefficiency. Economists have studied a wide variety of tools that can induce desired behaviour and align interests. These range from pay-for-performance to softer factors such reputation and norms. Each tool has strengths and weaknesses making them suitable to different environments.

Overview of the thesis

This thesis is chiefly an inquiry into the power of incentives to shape behaviour. Each chapter contributes to this inquiry in a different way, differing in the shape the incentive takes, the environment it is employed in and the research method employed. Chapters 2 and 3 employ game theory to analyze scenarios in which the measurement of performance is imperfect. Chapters 4 and 5 employ field experiments in order to investigate empirically how people respond to incentives.

Chapters 3 and 4 deal both with multitasking and the involvement of man-

agers. In both chapters agents (workers) perform multiple tasks while the performance measure computes a single outcome. This can lead to a misallocation of efforts if measured performance is not equal to the actually desired outcome. In chapter 3 we show that a manager with his or her own preference for the agent's execution of tasks can alleviate this if given some power over compensation. In chapter 4 we conjecture that task assignment in a retail team is not solely aimed at maximizing revenue but is also used for other purposes such as training and maintaining a pleasant workplace atmosphere. We suggest that a team incentive can help focus teams on sales maximization by inducing harder and smarter work through task assignments, but find no evidence for either. The multitasking problem is less present in chapters 2 and 5. The students of chapter 5 face a multitasking problem in that they have another course to complete, while chapter 2 contains an element of multitasking in that a worker spreads efforts over time rather than over tasks. Note also that while the mentors of the students in chapter 5 are not supervisors in the strict sense they do fill a supervisory role in that they check with students regarding their study performance.

In chapters 2, 3 and 5 the incentives are based solely on the agent's own performance. These chapters mostly employ goals or targets (chapter 2 also considers linear pay-for-performance). In chapters 2 and 3 meeting targets results in an extrinsic reward while in chapter 5 the study goal is used to motivate students without an extrinsic reward. Chapter 4 meanwhile considers a tournament based incentive where teams win a prize depending on their own performance and that of their peer group.

Thus the main themes of this thesis are performance (mis)measurement and the utilization of performance measures in incentive contracts, the role of supervisors and the distribution of efforts over multiple tasks. The remainder of this section discusses the chapters in more detail.

In chapter 2 I study the optimal incentive contracts when a performance measure is used that is less granular than the underlying production process. Specifically I consider the scenario in which the performance measure cannot

Introduction

distinguish between efforts provided before and after more information regarding the value of efforts becomes available. Changes in information regarding the production function should optimally result in efforts being adjusted as well, whether upwards or downwards. However, a change in efforts is potentially hampered by a contract based on a sticky performance measure. A simple two-period principal-agent model is employed to study the scenario.

In the chapter I show that the problem of properly incentivizing the agent using the imperfect performance measure can be overcome through proper contract design. More specifically if the imperfect performance measure is output-based initial contracts can be set up that provide the agent with the proper incentive to respond regardless of the signal received mid-contract. In contrast, when the performance measure is based on inputs additional contracting occurs as information is revealed. This renegotiation shapes the form of the initial contract, which must be set up to maximize value in either of the extremes of the distribution of the value of efforts. Such an extreme contract reduces the bargaining power of the agent by setting up the contract in such a way that renegotiation is either always concerned with lowering or always with raising efforts (depending on the extreme chosen). Thus regardless of whether an input-based or output-based performance measure is used a contractual form can be found that results in the proper adjustment of efforts to new information. Note however that in the output-based scenario the agent drives the response to new information, such that this scenario is preferred in case the agent receives private information regarding the value of efforts.

In chapter 3 Josse Delfgaauw and I study the use of subjective and objective performance pay in a setting where an agent performs multiple tasks. Pay-for-performance based on an objective performance measure has the advantage of being easily contractible, but may provide an incorrect impression of actual performance. In such cases subjective pay-for-performance administered by a middle manager can be more accurate. However this need not be the case if managers have their own agenda. We study the interplay between principal, supervisor and

agent in a multitasking setting with an imperfect aggregate performance measure and biased supervision.

By determining what part of the agent's compensation the supervisor has control over and the conditions imposed on pay based on the objective performance measure the principal can control the supervisor's authority over the agent. This allows the principal to use the supervisor as an indirect tool to control the agent. We show that a supervisor with a preference over the tasks has the same distorting effect on the agent's efforts as an equally incongruent verifiable performance measure. However, we further show that biased supervision is not always a perfect substitute for an incongruent objective performance measure. Whereas multiple incongruent performance measures can be combined to increase the alignment of an incentive system (e.g. Feltham and Xie 1994, Datar et al. 2001, Budde 2007) we show that adding a supervisor with discretionary power over compensation can but need not increase congruence. This is because the supervisor can change her performance report depending on how such reports are being interpreted by the principal.

Chapter 4 contains a field experiment on the effects of team incentives on sales and task assignment in a retail setting joint with Robert Dur and Josse Delfgaauw. Team performance commonly depends not just on the efforts of the individual team members but also on task assignment. After all, specialization and the exploitation of comparative advantage is one of the cornerstones of economics. However, when assigning tasks to team members performance may not be the only aspect taken into account. Favouritism, employee seniority, employee preferences, and fairness considerations can also play a role. By increasing the importance of team performance for all team members a team incentive can curtail these factors in the task assignment decisions, in addition to eliciting stronger team effort.

To investigate whether and to what extent team incentives improve performance and influence task assignment we introduced a team incentive in a random subset of 108 stores of a retail chain. The incentive consisted of a six week com-

Introduction

petition among the stores, each store being paired with two similar stores. Each employee of winning stores received a bonus of roughly three percent of monthly earnings. Using surveys before and directly after the incentive period we investigate the effect on task assignment and job satisfaction. We find no effect of the team incentive on either performance or task assignment. Our estimate regarding performance is fairly precisely estimated, suggesting a genuine non-effect rather than a lack of statistical power. While we expected ability to become more important in task assignment the survey results show no change in its importance for the assignment tasks. Together these results imply that the incentive scheme did not affect employee behaviour.

In chapter 5 we move away from employees, as Max van Lent and I study the effects of motivating students to set a grade goal on study performance in a field experiment among almost 1100 first-year economics students. As discussed earlier goals can provide a reference point that helps motivate. We are interested not only in whether self-set goals can help improve study performance but also whether challenging students to set more ambitious goals leads to additional improvement. We expect that if student's self-set goals are of moderate or easy difficulty, being prompted to raise the goal can improve performance by making the goal harder to meet but still attainable.

The first-year students were randomly assigned to the control, goal or raise (challenge) group. All students regularly have meetings with their university assigned mentor. In the goal treatment the mentor was tasked with asking students whether they had a specific grade goal in mind for the main course of the period, while in the raise treatment the mentor was additionally tasked with challenging students to raise their goal if she thought the goal was of moderate or easy difficulty. We find that students assigned to the goal treatment perform roughly 9% of a standard deviation better than students in the control group, an effect size in line with that of other interventions in education (Sanders and Chonaire 2015). The result is driven by a lower drop out rate. Students in the raise treatment however are found to have results similar to those in the control group, suggesting

that challenging students to raise their goals is counterproductive.

The purpose of this thesis is to shed light on how incentives work. Chapters 2 and 3 utilize game theory to provide clear predictions regarding what incentive contracts should look like if the premises in those chapters hold. Chapters 4 and 5 put theory to the test by means of field experiments to see how incentives operate in the real world. As stated earlier the overarching theme of this thesis is that incentivizing others is possible but not necessarily easy. The experimental results show that incentives can but need not have the effect predicted by theory. This shows the importance of further enlarging our understanding of when what incentive mechanism works.

Chapter 1

Incentives and Limited Measurement

1.1 Introduction

When two parties engage in a contract they are essentially performing a balancing act. The contract is designed to implement a goal, but is restricted to information pertaining to that goal that is known or imaginable at the time of writing. When circumstances change, the performances contracted upon may no longer be desirable, and hence the contract unfit for its purpose. For instance the marketing manager of a firm may be assigned the task of working a product market. If the response to these efforts is lower than expected it becomes unproductive to stick to the efforts as agreed upon. More generally, parties contracting upon a prolonged effort that cannot be broken up into separate contracts may find that information regarding the value of the efforts changes during the contract.

A dynamic environment makes adapting actions to the environment valuable. At the moment of signing a contract the contractee is limiting her ability to respond to a changing environment. This paper addresses how the principal should deal with contracting on measured performance when the value of that

performance is uncertain *ex ante*.

The question posed in this paper is relevant when the value of efforts (i) cannot be contracted upon and (ii) is initially unknown, but (iii) efforts can be changed during the contract even though (iv) they cannot be contracted upon separately before and after the revelation of information regarding their value (but renegotiation is possible). There are many reasons why the value of efforts may at first be unknown. Actions of competitors, market size, consumer preferences, efficiency gains from specialization or scale, and the progress of other agents are all factors that can be estimated but not known with certainty. The value of efforts is revealed as these factors become known. Depending on the the extent of the uncertainty regarding the value of efforts it can be costly to write a full contract stipulating required performance under all possible circumstances, and simply impossible if the value of efforts is unverifiable. At the same time it can be impossible to break up the contract into smaller contracts. In particular if the performance measure(s) relied upon are not granular enough to provide information regarding performance prior to and subsequent to information revelation. This could be because the performance measure is imperfect and a function of all inputs (eg. the efforts over time build up to a single outcome, or initial efforts are complementary to later efforts), or the principal has found reason not to invest in setting up a detailed performance measurement system (eg. due to the costs involved).

In this paper we develop a simple two period principal-agent model to analyze the situation described above. The principal and agent contract at the start of the game using a performance measure that is a function of first and second period efforts. Measured performance realizes only at the end of the second period. Information regarding the value of efforts is revealed at the start of the second period to all, allowing the agent to tailor second period efforts to the actual rather than expected value of his efforts. The agent may optimally have to provide more or less effort than anticipated initially. In solving the game we consider both the use of linear (piece-rates) and non-linear (target based) pay-for-performance.

Introduction

We find that using an output-based performance measure does not require the principal to react to the revelation of information. Using non-linear, target based pay, she sets multiple performance targets with associated bonuses. This lets the agent use his information regarding the value of efforts to select which option to pursue. Using linear piece-rates she essentially sells the firm to the agent. Both approaches do not require awaiting the revelation of information. This stands in contrast to the case when an input-based performance measure is used. The principal then initially offers a contract that maximizes value in one of the extremes of the value of effort distributions, over- or underincentivizing the agent depending on the true value of effort. As information is revealed the principal renegotiates the contract (in case of linear piece-rates) or (unilaterally) offers an additional performance target with associated bonus (under target based pay). While piece-rate pay requires explicit consent, target based pay requires simply that the secondary offer is attractive enough to the agent. In both cases the initial contract protects the agent from extortion by the principal in the second period, thus preserving first period effort incentives.

The key findings of this paper are as follows. First, we show that regardless of the type of performance measure used first-best efforts can be elicited by the principal. The methods by which this optimal outcome is achieved differ. When using an output-based measure, the principal relies on the agent to respond to new information, whereas an input-based measure requires the principal to respond to new information by offering a new contract. Thus when information is not symmetric it is best to utilize an output-based measure if the agent is more informed and vice versa. Regardless, the imperfection in the performance measure, limited granularity, can be overcome through proper contract design. Second, this is true indiscriminately of the type of performance pay used. Piece-rates have the advantage of using a simple contract when an output-based performance measure is in use, but possible disadvantage of requiring the principal to renegotiate with the agent when an input-based measure is used. In contrast target-based pay requires the principal to set up a (potentially) elaborate contract in the output-based sce-

nario but allows unilateral offers to the agent. Finally, under input-based pay the initial contract will maximize value for one of the extremes of the value of efforts distribution. That is, the principal either sets up a contract that maximizes value if value of effort is as low as can be conceived or that maximizes value if effort is as high as it can conceivably be. As a result the contract is always adjusted in one direction, if at all. A principal expecting to be able to adjust a contract in two directions is easily taken advantage of by an agent who banks on renegotiating in one direction only. It is to preempt this that the initial contract is such that value is maximized in one of the extremes of the value of efforts distribution.

The paper is organized as follows. The related literature is discussed in Section 1.2. This is followed by the setup of the model in Section 1.3, after which we analyze the first-best benchmark case in Section 1.4 and the full model in Section 1.5. Section 1.6 concludes.

1.2 Related literature

This paper is closely related to the general literature on incentives and performance measures, and the literature regarding the use of input- versus output-based performance measures in particular. In a seminal paper Holmström (1979) shows that all measures that provide incremental information should be used in a contract. Following the results of Shavell (1979) on optimal risk-sharing between those engaged in a contract it has been pointed out that input based performance measures may be less impacted by factors not under the control of the agent. Thus input-based performance measure may weigh more in the optimal contract than output-based measures in order to lower the risk borne by risk averse agents. While risk-aversion of the agent may make it appealing to reward inputs, inputs can be costly to observe as pointed out by Lazear (1986). Furthermore Prendergast (2002) and Raith (2008) argue that the use of input-based performance measures requires that the principal knows how inputs translate into

Related literature

output. If the agent has superior information then a greater weight for output based measures may be required in order to induce the agent to efficiently use his private information. In this paper too the principal relies on the agent to use his private information when using an output-based performance measure but utilizes her own when an input-based measure is used.

The paper is also related to the literature on adaptation and authority going back to the seminal paper by Simon (1951), see e.g. Aghion and Tirole (1997), Dessein and Santos (2006), Hart and Holmström (2010), and Gibbons (2005) for an overview. This paper primarily shares the ex-ante uncertainty regarding the pay-offs of actions such that the optimal action(s) cannot be contracted upon beforehand. However it is not concerned with the assignment of decision rights, where the holder of the rights can (to a large extent) dictate the decision(s) made ex-post. Instead we focus on incentivizing the agent's ex-post to carry out the desired action. As shown in the paper the agent's actions can be optimally adjusted to the environment but this requires either a potentially elaborate contract up front (output-based performance pay) or renegotiation of the initial contract (input-based performance pay), the latter of which is generally assumed to be infeasible in this literature. However, using an output-based performance measure effectively amounts to assigning the agent the decision rights regarding effort, as he is relied upon to respond to new information (more so if the information is private).

Another strand of related literature, especially for the case of input-based performance measures, is that on ratchet effects. Ratcheting is a potential concern when future incentive contracts may change based on past performance, see eg. Weitzman (1980), Freixas, Guesnerie and Tirole (1985) and Gibbons (1987). Intuitively, the principal can use past performance to deduce characteristics of the production function. Hence, high past performance may be used to ratchet up future performance demands, bringing demands more in line with the actual potential of the business. There is a small literature documenting such adjustments, see e.g. Bol and Lill (2015), Arnold and Artz (2015) and Aranda et al.

(2017). Arnold and Artz (2015) show initial target difficulty makes target flexibility more likely. Studying the evolution of targets at new branches of a travel organization Aranda et al. (2017) show past performance is an increasingly important factor in setting performance targets. However, ratcheting potentially destroys the incentive to perform in the present through the anticipation of adjustments. Bouwens and Kroos (2011) indeed show that retail store managers attenuate performance in the last quarter of the year in order to lower future target increases. Bol and Lill (2015) show that firms may be hesitant to ratchet, limiting target adjustments unless the circumstances (the economics of the business) have changed, allowing the preservation of incentives. In this paper the principal may offer a new contract to the agent as the value of effort becomes known, but cannot change the initial contract based on past performance. She does so when using an input-based performance measure after both information regarding the value of efforts is revealed and the agent has put in some effort. The agent wants to anticipate on the secondary contract, for instance by initially exerting more or less effort than warranted by the first contract. The principal however will want to incorporate all efforts already provided into the second contract. If the principal succeeds in doing so the agent does not benefit from his anticipatory efforts. In order to maintain incentives for the agent the principal offers an initial contract that protects the agent from such exploitation. However, the initial contract must be aimed at maximizing value in one of the extremes of the distribution of the value of efforts. If not the agent can take advantage of the initial contract by selectively negotiating a new contract.

The potential exploitation of the agent by the principal in a renegotiation of the contract after the agent has already put in some efforts is reminiscent of the hold-up problem. The hold-up problem, as discussed in seminal contributions by Grossman and Hart (1986) and Hart and Moore (1990), occurs when parties must make a relationship-specific investment for a transaction but the optimal transaction cannot be contracted upon beforehand. In this case the principal and agent cannot contract upon the value of effort as it is unverifiable. As

the principal can hold up the agent by skimming off of all possible rents in a renegotiated contract the principal must offer an initial contract to ensure the agent is compensated for his efforts before the value of efforts becomes known. The agent in the second period can also hold up the principal in the second period by working to the letter of the initial contract. In essence, the initial contract provides the agent a fallback option that not only protects the agent's remuneration for expected first period efforts, but also offers him a secure position from which to gamble upon whether effort will be more or less valuable than initially expected. By working too little or too much in the first period the agent can preselect into when to enter into negotiations with the principal when the value of efforts becomes known. The optionality offered by the existence of the first contract forces the principal to offer an initial contract focused on maximizing value in one of the possible extremes of the value of efforts, such that the agent is restricted in his speculation. Allowing the principal to offer the amendment to the contract is tantamount to giving her all bargaining power. As will be shown this is natural in case of performance targets where the principal can unilaterally announce an additional performance target and associated bonus, but less obvious in case of piece-rate pay since it requires renegotiation implying the agent's consent.

One case considered in this paper is the use of a renegotiated linear piece-rate when an input based performance measure is utilized. This is because the uncertainty regarding the value of effort makes for a situation in which it is ex-ante not known what a non-linear piece-rate scheme should look like. The literature points to another reason why linear piece-rates are commonly used. Linear piece-rate contracts exhibit robustness, protecting the principal from gaming by the agent since the principal always obtains a given cut of each marginal unit of measured performance. For instance in Holmström and Milgrom (1987) the agent performs tasks over time and show that it is optimal to use a contract that is linear in an aggregated performance measure as the agent could adapt the sequence of actions in response to realized results if a non-linear pay-for-performance scheme

is used. Note that in contrast to this paper the environment in Holmström and Milgrom (1987) is time-independent such that the agent's optimal effort remains constant over time. Carroll (2015) shows that a linear contract protects the principal if the agent can take actions unknown to the principal. Note that the use of targets in this paper also allow the principal to implement non-linear compensation for efforts while guaranteeing her a certain outcome.

1.3 The model

We consider a principal-agent model in which all players are risk neutral. The total wage payment to the agent is w^A , and his outside option utility equals zero.

The game consists of two periods $t \in \{1, 2\}$ during which the agent works on a single task. Cost of effort in a single period is $c(e_t) = \frac{1}{2}e_t^2$. Thus the agent's utility is given by:

$$U^A = w^A - \frac{1}{2}e_1^2 - \frac{1}{2}e_2^2 \quad (1.1)$$

The principal's utility is given by:

$$U^P = v(e_1 + e_2) - w^A \quad (1.2)$$

where v is the marginal value of effort to the principal. Note that effort is equally valuable to the principal regardless of the period. The value of effort is unverifiable and furthermore initially unknown, being revealed only at the start of t_2 . For simplicity we let $v \in \{l, h\}$, with $Pr(v = h) = \frac{1}{2}$ and $l < h$.

The agent's efforts are private information of the agent but the principal has access to a verifiable performance measure $m(e_1, e_2, v)$ that can be used as a basis for an incentive contract. An output (input) based performance measure is (in)dependent of v . For simplicity take $m^I = e_1 + e_2$ to be an input based performance measure and $m^O = v(e_1 + e_2)$ to be the output based performance measure under consideration.

The principal and agent agree to a contract at the start of the first period

before v is known. After the realization of v becomes at the start of the second period they can renegotiate, or the principal can unilaterally offer the agent additional compensation. We assume that the principal cannot commit to offering a certain contract at t_2 because the value of effort v is unverifiable even after it has been observed by the principal and agent.¹

We consider two specific contractual forms, non-linear performance pay using targets and a linear piece-rate. Under piece-rate pay the agent receives a wage consisting of a fixed portion a and a bonus b per unit of measured performance, i.e. $w^A = a + bm$. If the principal wishes to change the conditions of the contract a or b she will have to renegotiate with the agent. The non-linear payment scheme considered consists of target-based pay where the agent earns total compensation b^i when meeting or exceeding a target level of measured performance $m = m_i$. This allows the principal to offer a menu at the start of the game and to offer a contract amendment by stipulating additional performance levels and associated bonuses at the start of the second period. Thus the principal can offer a target m_1 with associated bonus b_1 at the start of the first period and offer another target with associated bonus b_2 at the start of the second period. Note that while the first contract (m_1, b_1) must induce the agent to accept the contract, the second need only be incentive compatible in order to induce the agent to respond. Setting $m_2 > m_1$ ($m_2 < m_1$) with proper associated bonus b_2 allows the principal to induce the agent to raise (lower) second period effort.

The timing of the game is as follows:

1. Nature draws $v \in \{l, h\}$.
2. The principal offers the agent a contract, determining w^A .
3. The agent accepts or rejects the contract. If the agent rejects, the agent and principal receive their outside option payoff.
4. The agent chooses first period effort e_1 .

¹In addition, the multiplicity of possible values of v may simply make it unfeasible to contract on v beforehand.

5. The value of effort v is revealed.
6. The principal may offer an amendment to the contract to the agent, which the agent can accept or reject.
7. Measured performance m is realized and players receive their payoff.

We use backward induction to solve for a sub-game perfect Nash equilibrium.

1.4 First-best Benchmark

Suppose that efforts can be contracted on directly such that there is no friction in contracting between the principal and agent. The effort levels that maximize joint welfare are found by maximizing the total welfare functions at time $t = 1$ (for e_1) and $t = 2$ (for e_2):

$$\begin{aligned}\max_{e_1} E[TWF] &= E[v] e_1 - \frac{1}{2} e_1^2 \\ \max_{e_2} E[TWF] &= v e_2 - \frac{1}{2} e_2^2\end{aligned}$$

where $E[v] = \frac{1}{2}(l + h)$, the expected value of efforts at t_1 . The resulting first order conditions solve for:

$$\begin{aligned}e_1^{FB} &= \frac{1}{2}(l + h) \\ e_2^{FB} &= v\end{aligned}$$

Hence the agent should exert more effort in the second period if value is demonstrably high and less if it is low. In the first period there is uncertainty on the value of effort which is reflected in the first-best first period effort level. The agent optimally puts in a first period effort level that is a weighted average (reflecting the probabilities of the various possible levels of v), since too little effort is a lost opportunity in terms of value creation if value turns out to be high but too much means that the value created is being eroded by too high

first period effort costs. This results in $E[TWF] = \frac{1}{8}(l+h)^2 + \frac{1}{4}l^2 + \frac{1}{4}h^2$, or $E[TWF] = \frac{1}{4}(\frac{3}{2}h^2 + hl + \frac{3}{2}l^2)$.

It should be noted that given that $e_2^{FB}(v=l) < e_1^{FB} < e_2^{FB}(v=h)$ the central problem of the principal is that she must be able to induce the agent to exert efforts in a flexible way. Given separable effort costs and absent the ability to commit on contracts based on the value of effort this requires that she must be able to pay less per unit of effort in the second period in order to induce the agent to exert lower efforts in the second period than he did in the first. Thus the situation under consideration lends itself to non-linear pay-for-performance contracts that enable the principal to do this.

1.5 Analysis

Consider now the game when the principal cannot contract on efforts directly nor commit based on the value of efforts v . The principal will offer an initial contract which can be renegotiated at the start of the second period.

We consider two cases, first turning to output-based measured performance before turning to input-based measured performance. For each type of performance measure a non-linear performance pay mechanism (target based pay) and linear performance pay is analyzed.

1.5.1 Output based measured performance

Target based pay

As discussed above the primary problem is that the agent optimally exerts different levels of effort in the two periods, lowering effort in the second period compared to the first depending on the value of effort v . We focus here on the case where measured performance is a function of output, hence $m^O = v(e_1 + e_2)$. Specifically we consider whether the principal can offer measured output-bonus combinations (m_h^O, b_h) and (m_l^O, b_l) at the start of the game that induce opti-

mal efforts from the agent. Thus the principal offers multiple sets of measured performance and associated pay. The results are given in Proposition 1 below.

Proposition 1 *In case of an output-based performance measure and target based pay the principal implements the first-best outcome by offering two measured (output) performance targets and associated bonuses combinations (m_i^O, b_i) at the start of the game. The optimal target-bonus sets are $m_l^O = \frac{3}{2}l^2 + \frac{1}{2}lh$ with $b_l = \frac{3}{16}h^2 + \frac{1}{4}hl + \frac{9}{16h^2}l^4$ and $m_h^O = \frac{3}{2}h^2 + \frac{1}{2}hl$ with $b_h = \frac{1}{4}hl - \frac{9}{16h^2}l^4 + \frac{9}{16}h^2 + \frac{3}{4}l^2$. This results in first-best efforts, $e_1 = e_1^{FB}$ and $e_2 = e_2^{FB}$.*

Proof. The proof is given in the appendix. ■

Proposition 1 shows that output based pay-for-performance allows the principal to commit to a payment scheme. This commitment then empowers the agent to optimally react to the resolution of uncertainty regarding the value of efforts. The agent observes the targets and associated bonuses and decides which to pursue depending on the observed realization of v at the start of the second period. Anticipating this delayed choice he puts in a moderate amount of effort in the first period.

Piece-rate pay

We now consider the case in which the principal utilizes a linear piece-rate and a fixed wage to incentivize the agent, $w^A = a + bm^O$. While the contract could be renegotiated following the revelation of v at the start of the second period this is not required as shown in Proposition (2).

Proposition 2 *In case of an output-based performance measure and piece-rate pay the principal optimally sets $a = -\left(\frac{3}{8}h^2 + \frac{1}{4}hl + \frac{3}{8}l^2\right)$ and $b = 1$. No renegotiation takes place. This results in first-best efforts, $e_1 = e_1^{FB}$ and $e_2 = e_2^{FB}$.*

Proof. The proof is given in the appendix. ■

Since the output-based performance measure tracks the value of the firm the principal can sell the firm to the agent who then internalizes the costs and

benefits. As under target based pay the agent weighs the odds of low and high value in the first period and reacts to the revelation of the value autonomously in the second period.

1.5.2 Input based measured performance

When the principal relies on an input-based performance measure, here $m^I = e_1 + e_2$, the primary problem is that for an input-based performance measure every unit of input impacts the measure equally but a unit of effort in the first period may be substantially more or less valuable to the principal compared to a unit of effort in the second period due to the revelation of information at the start of $t = 2$. When an output-based performance measure is used this discrepancy between the first and the second period is automatically taken care of since the agent uses his knowledge of the value of effort in setting efforts optimally. Instead, when utilizing an input-based performance measure the principal will have to set consecutive contracts in order to cope with this problem as we show below.

Target based pay

If the principal utilizes target based pay she can introduce a second target \underline{m}_2 , with associated total compensation b_2 if $m \geq \underline{m}_2$, at the start of the second period in order to react to v . She can do so unilaterally without explicit consent of the agent but is nevertheless bounded by what the agent is willing to do if such a contract is to have an effect. For ease of exposition we treat the second period bonus b_2 as an addition to the first period bonus b_1 (associated with $m \geq \underline{m}_1$), such that total compensation if \underline{m}_2 is met is $w^A(m \geq \underline{m}_2) = b_1 + b_2$. Note that we allow any \underline{m}_2 , including $\underline{m}_2 < \underline{m}_1$. Adding an additional performance target and bonus allows the principal to pay a lump sum for a given performance, effectively allowing a non-linear compensation per unit. However, the fact that the second performance target and associated bonus is set at the start of the second period has a potential drawback. If the initial performance target is too low it is in the

principal's interest that the agent works 'ahead of schedule' in the first period in order to save on effort costs. The agent may do so in anticipation of a possible higher performance target (and bonus) in the second period (eg. as would happen if the principal could commit). However, if the principal deduces this she will take advantage of the agent by setting a correspondingly higher goal or lower bonus in the second period. This, in turn, can deter the agent from spreading effort costs optimally over the two periods, ultimately resulting in higher implementation costs for the principal who is the final bearer of these costs. Since the principal cannot commit to leaving the agent a rent in the second period she can only overcome this by setting the first period target and bonus high enough such that the agent is not working ahead of schedule without compensation.

At the start of t_2 the agent chooses between meeting \underline{m}_2 and meeting \underline{m}_1 .² In either case the agent is best off not exceeding the target of choice in expectation, hence setting $e_2 = \underline{m}_t - e_1$. The agent is better off pursuing \underline{m}_2 if the difference in rewards b_2 exceeds the difference in effort costs:

$$\begin{aligned} U^{A,\underline{m}_2} &= b_1 + b_2 - \frac{1}{2}(\underline{m}_2 - e_1)^2 - \frac{1}{2}e_1^2 \geq b_1 - \frac{1}{2}(\underline{m}_1 - e_1)^2 - \frac{1}{2}e_1^2 = U^{A,\underline{m}_1} \\ b_2 &\geq \frac{1}{2}(\underline{m}_2 - e_1)^2 - \frac{1}{2}(\underline{m}_1 - e_1)^2 & (t_2 \text{ ICC}) \\ b_2 &\geq \frac{1}{2}(\underline{m}_2 - \underline{m}_1)(\underline{m}_1 + \underline{m}_2 - 2e_1) \end{aligned}$$

At the start of $t = 2$ the principal learns the value of effort v and can choose to offer the agent an additional measured performance target and associated bonus. It is in her interest to induce the agent to meet that target, hence her problem

²Here we assume that meeting \underline{m}_1 is preferred over meeting neither. This assumption holds in equilibrium.

is:

$$\begin{aligned}
 \max_{\underline{m}_2, b_2} E[U^P] &= v(\underline{m}_2 - E^P[e_1] + E^P[e_1]) - b_1 - b_2 \text{ s.t. } (t_2 \text{ ICC}) \\
 \max_{\underline{m}_2, b_2} E[U^P] &= v\underline{m}_2 - b_1 - \frac{1}{2}(\underline{m}_2 - E^P[e_1])^2 + \frac{1}{2}(\underline{m}_1 - E^P[e_1])^2 \\
 \frac{\partial E[U^P]}{\partial \underline{m}_2} &= v - (\underline{m}_2 - E^P[e_1]) \iff \underline{m}_2 = E^P[e_1] + v
 \end{aligned}$$

where $E^P[e_1]$ is the principal's belief regarding the agent's first period effort. In setting the new performance target the principal optimally takes into account that measured performance is to some extent due to efforts that have already been provided and as such are sunk to the agent, no longer influencing his second period decision. Thus the principal only pays for measured performance above and beyond first period expected measured performance. Note also that the principal optimally demands the efficient level of additional effort in the second period. Thus as long as $(t_2 \text{ ICC})$ holds and the principal's belief $E^P[e_1]$ is correct the agent supplies first-best second period effort.

When the agent is deciding on e_1 he deduces the $\{\underline{m}_2, b_2\}$ package the principal potentially offers in the second period and takes that into account. There are four scenarios that the agent's choice of effort at $t = 1$ can lead to: (i) he cannot be induced to pursue \underline{m}_2 , (ii) he can always be induced to pursue \underline{m}_2 , (iii) he can only be induced to pursue \underline{m}_2 if $v = l$, (iv) he can only be induced to pursue \underline{m}_2 if $v = h$. Note that under each scenario we assume that it is incentive compatible for the agent to exert effort in the first period, if only in order to eventually meet \underline{m}_1 . The result is given in the following proposition:

Proposition 3 *Using targets based pay in combination with an input-based performance measure there are two equilibria:*

(i) *The principal sets an initial target of $\underline{m}_1 = \frac{1}{2}(3h + l)$ with associated bonus $b_1 = \frac{1}{4}l(h + 3l)$. At $t = 2$ the principal offers $\underline{m}_2^l = \frac{1}{2}(h + 3l)$ with associated total compensation $b_1 + b_2^l = \frac{1}{4}l(h + 3l) + \frac{1}{2}(l^2 - h^2)$ if $v = l$, and no new contract if $v = h$. This results in first-best efforts, $e_1 = e_1^{FB}$ and $e_2 = e_2^{FB}$.*

(ii) The principal sets an initial target of $\underline{m}_1 = \frac{1}{2}(h + 3l)$ with associated bonus $b_1 = \frac{1}{4}h(3h + l)$. At $t = 2$ the principal offers $\underline{m}_2^h = \frac{1}{2}(3h + l)$ with associated total compensation $b_1 + b_2^h = \frac{1}{4}h(3h + l) + \frac{1}{2}(h^2 - l^2)$ if $v = h$, and no new contract if $v = l$. This results in first-best efforts, $e_1 = e_1^{FB}$ and $e_2 = e_2^{FB}$.

Proof. The proof is given in the appendix. ■

Proposition (3) states that there are two possible equilibria, in each of which the initial contract is aimed at achieving the optimal outcome in one of the possible (extreme) value realizations. Obviously the principal need not offer a new contract if the value realization the initial contract was tailored towards is realized. There are two striking features of the proposed equilibria. First of all the initial contracts are tailored towards maximizing value in either of the possible value realizations but nevertheless still yield the first-best outcome. The underlying reason for this is that the agent anticipates upon the fact that the initial contract can be amended. It is always in the agent's interest to meet the final target he chooses to accept at minimal effort costs or, in our model of separable convex effort costs, to spread efforts over the two periods as evenly as possible. Thus the agent must not over- or underexert himself in the first period, since overexertion in anticipation of the high value scenario is costly if the value of effort turns out to be low and vice versa. Thus while the initial contracts are tailored towards one of the extremes of the value of effort the agent does not exclusively tailor his first period efforts toward this outcome as he anticipates that that outcome is a possibility, not a fact.

The second striking feature is that the first-best outcome cannot be achieved through a 'middle-of-the-road' initial contract. An initial contract that is not tailored towards the low or high value of effort scenario is subject to gaming by the agent. As the agent optimally meets the final target he pursues at minimal effort costs it is in his interest that the targets he ultimately chooses to pursue are as close together as possible, such that first period effort can be as efficiently

as possible. Therefore it is in the agent's best interest to stick to the initial contract in one value scenario and accept an amended contract only in the other. In effect this allows the agent to presort into the possibility of a value realization and better spread his effort costs. For instance say the agent anticipates upon $v = h$. By increasing effort in the first period he lowers his total effort costs of meeting the expected \underline{m}_2^h , thus earning a rent in that scenario. If $v = l$ this would be very costly if the agent pursued a low \underline{m}_2^l , since a low target does not require as much total effort and hence necessitate very unbalanced efforts from the agent. However the agent has the option to pursue \underline{m}_1 instead. Thus the presence of the initial middle-of-the-road contract allows the agent to gamble upon an extreme value outcome in the first period by exerting higher or lower than first-best effort and sticking to the middle-of-the-road contract in case his gamble turns out to be off. This gamble opportunity exists as long as the initial contract is not tailored towards either of the extremes of v . Note that a middle-of-the-road contract is optimally always adjusted in the second period. However this requires a valuable enough initial contract in order to insure the agent from extortion by the principal in the second period bargaining since at that point the principal has an incentive not to reward the agent for effort already put in. Thus an initial contract must exist and the agent's gaming behaviour in anticipation of either of the value outcomes forces the principal to choose an initial contract that is tailored towards one of the value extremes. Note that since this gaming behaviour relies on exploiting the principal's expectation of first period effort a 'middle-of-the-road' contract is feasible if first period effort is observable to the principal.

Piece-rate pay

Using piece-rate pay the principal still faces the issue that while first and second period effort impact the performance measure equally they are not equally valued by the principal. As a result the principal will want to offer a different piece-rate for second period effort but cannot know what level that piece rate should

be. Since an optimal piece rate scheme sets a piece-rate in conjunction with a negative base wage to recoup any rents given to the agent the principal must renegotiate with the agent. This requirement of explicit consent contrasts with target based pay where the principal can unilaterally announce an addition to the existing contract.

In this paper we consider the case in which the principal fully renegotiates the wage scheme with the agent, denoting the initial contract $w_1^A = a_1 + b_1 m$ and the second contract $w_2^A = a_2 + b_2 m$. A partial initial contract, e.g. where the agent receives b_1 up to a certain level of $\bar{m} = \underline{m}$, could be considered but in practice such a contract is similar to a fully renegotiated contract. The reason is that the agent will always want to distribute efforts across the periods in an attempt to obtain a rent. Thus to induce the agent to provide enough efforts in the first period the principal ends up setting \underline{m} so that it covers efforts in both periods. Such partial contracts are thus similar to renegotiating the entire contract.

When renegotiating with the agent the principal must preserve any rents the agent obtains under the initial contract. The resulting contracts are given in the proposition below.

Proposition 4 *Using piece-rates with an input-based performance measure there are two equilibria:*

(i) *The principal sets an initial piece rate $b_1 = h$ with fixed wage $a_1 = -\frac{7}{8}h^2 - \frac{1}{4}hl - \frac{1}{8}l^2 + \frac{1}{4}l$. At $t = 2$ the contract is renegotiated if $v = l$, in which case $b_2^l = l$ and $a_2^l = \frac{1}{8}h^2 - \frac{1}{4}hl + \frac{1}{4}l - \frac{9}{8}l^2$. No renegotiation takes place if $v = h$. This results in first-best efforts, $e_1 = e_1^{FB}$ and $e_2 = e_2^{FB}$.*

(ii) *The principal sets an initial piece rate $b_1 = l$ with fixed wage $a_1 = -\frac{7}{8}h^2 - \frac{1}{4}hl - \frac{1}{8}l^2 + \frac{1}{4}l$. At $t = 2$ the contract is renegotiated if $v = h$, in which case $b_2^h = h$ and $a_2^h = -\frac{9}{8}h^2 - \frac{1}{4}hl + \frac{1}{4}h + \frac{1}{8}l^2$. No renegotiation takes place if $v = l$. In both equilibria the first-best outcome is achieved with $e_1 = e_1^{FB}$ and $e_2 = e_2^{FB}$.*

Proof. The proof is given in the Appendix. ■

Note that during renegotiation the agent is reimbursed for any lost income on (expected) first period effort in addition to the difference in rents he can earn in the second period on the old contract compared to the new contract. Together with a binding participation constraint this ensures that the agent earns no rents. Similar to the result in Proposition (3) regarding the use of targets the initial contract offered is optimized for one of the extremes of v . Here too the reason is that if the principal does not do so she leaves a profitable gamble opportunity for the agent to focus on one of the value realizations. A naive principal is taken advantage off by the agent who gambles by for instance exerting high e_1 in anticipation of $v = h$. If the agent's gamble turns out correct he gets higher compensation after renegotiation than the principal expects in the form of a high piece rate over more units of performance. If the other scenario comes to pass the agent can fall back on the initial contract. This makes that he will only accept a new contract if it is closer to the first period contract compared to the optimal contract without gaming by the agent. In essence the initial contract offers the agent optionality when renegotiating. This is valuable to the agent but costly the principal as it can be used for gaming. The principal reduces the optionality of the initial contract by having it maximize total value for one of the extremes of the value of effort distribution. This ensures the agent can only game in one direction, known by the principal.

Although the outcomes are similar there are some noteworthy differences between the use of linear pay-for-performance (piece-rates) and that of non-linear pay-for-performance (through targets). First, the gaming of the agent takes place under both periods when targets are used whereas it only occurs in the first period when piece-rates are used. Targets require the agent to literally hit a target through a combination of first and second period efforts, which are thus inextricably linked. First period gaming thus affects second period effort when targets are used. Under piece-rates no such link exists as the agent is remunerated for his marginal productivity. Thus the piece-rate constitutes a very clear incentive for the agent that is not easily gamed in the second period. A final difference

between the types of pay to note is that the renegotiation required when linear pay-for-performance (piece-rates) is used requires the explicit consent of the agent, whereas in case of non-linear pay-for-performance (through targets) requires the implicit consent of the agent.

1.6 Discussion and concluding remarks

We have shown that if a principal and agent contract upon a task whose value is unknown but is revealed over time the first-best outcome can still be implemented even if performance measures cannot distinguish between the performance attributable to the period before and after the revelation of information. This is true regardless of whether input- or output-measures are used, even though the contracts to implement the first-best differ in form. Both non-linear performance pay, consisting of targets with an associated bonus, and linear piece-rate pay, where the agent receives compensation per unit of measured performance can yield the first-best outcome regardless of the type of performance measure.

Given information revelation during the game the optimal level of effort varies over time such that the principal optimally rewards (expected) units of performance in a non-linear fashion. The expected impact of efforts on output varies with the revelation of information. Therefore utilizing an output-based measure allows a single linear piece rate to accomplish non-linear compensation, and allows the creation of a menu of options for target-based pay. Either type of performance pay in combination with an output-based performance measure leads to contracting at the start of the game only. Thereafter the agent must respond to new information regarding the value of effort (productivity). In contrast, input-based measured performance requires some form of contract amendment. In the case of piece-rates the contract is renegotiated, whereas using targets the principal can single-handedly announce an addition to the existing contract even if she must respect the incentives of the agent in order for the addition to affect the agent's behaviour. Thus while a piece-rate contract is simpler when an output-based

Discussion and concluding remarks

performance measure is used, target based pay is more easily adjusted when necessary. The requisite adjustment of contracts when an input-based performance measure is used suggests that intermediate contracting is more common when a company utilizes input-based performance measures rather than output-based performance measures.

The necessity to renegotiate an input-based performance measure has strong implications for the initial contract. An initial contract provides a fallback position for the agent in future negotiation such that his compensation for efforts provided initially is guaranteed. However, this fallback position is a double-edged sword for the principal. On the one hand it allows her to make a credible commitment to the agent. On the other hand, the fallback position can be used by the agent in order to speculate upon the true value of efforts. For instance, he might exert little effort initially in hopes of the value of efforts turning out to be low. The costs to the agent of speculation are limited by his option of sticking to the initial contract in case he is wrong.³ An initial contract that allows the agent to speculate in either direction thus necessarily prevents adjustment of the agent's efforts in the second period in some scenarios. It is therefore best to limit the agent's ability to speculate to a single direction. Hence an initial contract that maximizes value in either of the extremes of the value of effort distribution is optimal. Note that while we have assumed there to be just two possible realizations of the value of efforts this logic holds even if there is a full distribution of the value of efforts.

This paper shows that if the nature of the task being contracted on is not fixed, contracts need to be accomodating through either increased contract complexity or intermediate adjustments. If the game were a stage game of a repeated game the principal and agent could instead form a relational contract, removing dependence on the relatively complex contracts. A relational contract can occur in a repeated game where the principal rewards the agent subjectively each pe-

³Essentially the fallback scenario of the initial contract allows the agent to compress the variation in efforts.

riod, being constrained in setting too low a bonus by losing future incentives for the agent but at the same time limited in setting high bonuses by her incentive to renege on payment, see Bull (1987) for an early treatment of such contracts and Malcomson (2013) for a recent overview.

The model assumed that the principal and agent have symmetric information regarding the value of effort. While the principal and agent must have compatible beliefs in order to agree to an initial contract they subsequently need not have the same information. With output-based performance measures the agent's beliefs regarding productivity determines how much effort he puts in, as he selects the output goal towards which he works. In contrast, when the performance measure is based on inputs the principal effectively decides upon efforts. With symmetric beliefs regarding productivity there is no difference between using an input- or output-based measure in terms of the outcome achieved. This changes however once the revelation of productivity is private information. For example, when only the agent receives information regarding his true productivity the use of an input-based performance measure is unable to implement the first-best since it relies upon the revelation of information to the principal. In contrast an output-based contract allows the agent to exercise good judgment and can implement the first-best. The reverse holds in case the principal receives private information. Prendergast (2002) and Raith (2008) similarly point out that output-based performance measures best incentivize the agent to utilize his private information, while input-based measures utilize the principal's information. Note further that this implies that when utilizing an input-based performance measure the principal has the strongest incentive to discover the value of effort while the agent has a strong incentive to learn the value of effort when an output based performance measure is used.

To conclude, we have shown that shortcomings in a performance measure, in the form of being unable distinguish between performance due to efforts before and after the true value of efforts becomes known, can be overcome through increases in contract complexity.

1.A Appendix

Proof of Proposition 1. The game is solved using backward induction. At $t = 2$ the agent observes v the actual value of his efforts and chooses between pursuing $m^O \geq \underline{m}_h^O$ or $m^O \geq \underline{m}_l^O$, where by assumption $\underline{m}_h^O > \underline{m}_l^O$. In order to meet any \underline{m}_v (dropping superscripts) the agent must put in $e_2 = \underline{m}_v/v - e_1$. Hence the incentive compatibility constraint (ICC) to choose \underline{m}_h over \underline{m}_l when $v = h$ is:

$$\begin{aligned} E[U^A|\underline{m}_h] &\geq E[U^A|\underline{m}_l] \\ b_h - \frac{1}{2} \left(\frac{\underline{m}_h}{h} - e_1 \right)^2 - \frac{1}{2} e_1^2 &\geq b_l - \frac{1}{2} \left(\frac{\underline{m}_l}{h} - e_1 \right)^2 - \frac{1}{2} e_1^2 \\ b_h - \frac{1}{2} \left(\frac{\underline{m}_h}{h} - e_1 \right)^2 &\geq b_l - \frac{1}{2} \left(\frac{\underline{m}_l}{h} - e_1 \right)^2 \\ b_h - b_l &\geq \frac{1}{2} \left(\frac{\underline{m}_h}{h} - e_1 \right)^2 - \frac{1}{2} \left(\frac{\underline{m}_l}{h} - e_1 \right)^2 \end{aligned} \quad (1.3)$$

Similarly the ICC to choose \underline{m}_l over \underline{m}_h when $v = l$ is:

$$\begin{aligned} E[U^A|\underline{m}_h] &\leq E[U^A|\underline{m}_l] \\ b_h - \frac{1}{2} \left(\frac{\underline{m}_h}{l} - e_1 \right)^2 - \frac{1}{2} e_1^2 &\leq b_l - \frac{1}{2} \left(\frac{\underline{m}_l}{l} - e_1 \right)^2 - \frac{1}{2} e_1^2 \\ b_h - \frac{1}{2} \left(\frac{\underline{m}_h}{l} - e_1 \right)^2 &\leq b_l - \frac{1}{2} \left(\frac{\underline{m}_l}{l} - e_1 \right)^2 \\ b_h - b_l &\leq \frac{1}{2} \left(\frac{\underline{m}_h}{l} - e_1 \right)^2 - \frac{1}{2} \left(\frac{\underline{m}_l}{l} - e_1 \right)^2 \end{aligned} \quad (1.4)$$

We will show below that these constraints hold and are non-conflicting.

Assume the agent pursues \underline{m}_v for each possible v such that $e_2 = \frac{\underline{m}_v}{v} - e_1$. In the first period the agent's problem is:

$$\max E[U^A] = \frac{1}{2} \left(b_h - \frac{1}{2} \left(\frac{\underline{m}_h}{h} - e_1 \right)^2 \right) + \frac{1}{2} \left(b_l - \frac{1}{2} \left(\frac{\underline{m}_l}{l} - e_1 \right)^2 \right) - \frac{1}{2} e_1^2$$

Resulting in:

$$e_1 = \frac{1}{4} \left(\frac{\underline{m}_h}{h} + \frac{\underline{m}_l}{l} \right) \quad (1.5)$$

Note that the outputs are deconstructed into the equivalent effort levels by weigh-

ing with the relevant value of effort v .

Substituting these results back into $E[U^A]$ gives:

$$E[U^A] = \frac{1}{2}b_l + \frac{1}{2}b_h - \frac{3}{16}\frac{m_h^2}{h^2} - \frac{3}{16}\frac{m_l^2}{l^2} + \frac{1}{8}\frac{m_h m_l}{hl}$$

The principal maximizes utility subject to the agent's participation constraint $E[U^A] \geq 0$, and the second period incentive compatibility constraints derived above. Let the participation constraint bind and substitute into $E[U^P] = \frac{1}{2}(\underline{m}_h + \underline{m}_l) - \frac{1}{2}(b_l + b_h)$ to obtain the principal's problem under the assumption that the incentive compatibility constraints hold:

$$\max_{m_h, m_l} E[U^P] = \frac{1}{2}(\underline{m}_h + \underline{m}_l) - \frac{3}{16}\frac{m_h^2}{h^2} - \frac{3}{16}\frac{m_l^2}{l^2} + \frac{1}{8}\frac{m_h m_l}{hl}$$

Yielding first order conditions $\underline{m}_h = \frac{4}{3}h^2 + \frac{1}{3}h\frac{m_l}{l}$ and $\underline{m}_l = \frac{4}{3}l^2 + \frac{1}{3}l\frac{m_h}{h}$ which solve for:

$$\begin{aligned}\underline{m}_h &= \frac{3}{2}h^2 + \frac{1}{2}hl \\ \underline{m}_l &= \frac{3}{2}l^2 + \frac{1}{2}lh\end{aligned}$$

which it should be noted are first-best output levels. The resulting effort levels are also first-best $e_1 = \frac{1}{2}(h + l)$, and $e_{2,h} = h$ and $e_{2,l} = l$.

What remains is to check the assumption that the incentive compatibility constraints hold. Utilizing the binding participation constraint and the results above allows simplifying the two constraints to:

$$\frac{3}{4}h^2 + \frac{1}{4}hl + \frac{9}{16}l^2 - \frac{9}{16}\frac{h^4}{l^2} \leq b_l \leq \frac{3}{16}h^2 + \frac{1}{4}hl + \frac{9}{16h^2}l^4$$

It can be easily shown that if one of these constraints is binding the other is slack, since if one binds the condition can be rewritten to $-\frac{9}{16h^2l^2}(h^2 - l^2)^2(h^2 + l^2) \leq 0$, which always holds. Letting (1.3) bind and substituting the result into the

binding participation constraint yields the reported bonus levels of:

$$\begin{aligned} b_l &= \frac{3}{16}h^2 + \frac{1}{4}hl + \frac{9}{16h^2}l^4 \\ b_h &= \frac{1}{4}hl - \frac{9}{16h^2}l^4 + \frac{9}{16}h^2 + \frac{3}{4}l^2 \end{aligned}$$

■

Proof of Proposition 2. Backward induction is used to solve the game. We show here that a single contract $w^A = a + bm$ set at the start of the game implements the first-best outcome.

At $t = 2$ the agent observes v and hence maximizes his utility $U^A = a + bv(e_1 + e_2) - \frac{1}{2}e_1^2 - \frac{1}{2}e_2^2$. From the first order condition $\frac{\partial U^A}{\partial e_2} = bv - e_2 = 0 \iff e_2 = bv$. At $t = 1$ the value of efforts is unknown and the agent maximizes $E[U^A] = a + \frac{1}{2} \left(bl(e_1 + e_2) - \frac{1}{2}e_{2,l}^2 \right) + \frac{1}{2} \left(bh(e_1 + e_2) - \frac{1}{2}e_{2,h}^2 \right) - \frac{1}{2}e_1^2$, where $e_{2,v} = bv$ as shown above. This results in $\frac{\partial U^A}{\partial e_1} = \frac{1}{2}b(l + h) - e_1 = 0 \iff e_1 = \frac{1}{2}b(l + h)$.

The principal at the start of the game sets a and b such as to maximize her utility $U^P = \frac{1}{2}(1 - b)l(e_1 + e_{2,l}) + \frac{1}{2}(1 - b)h(e_1 + e_{2,h}) - a$. Where e_1 and $e_{2,v}$ are subject to the constraints derived above. She is best off setting the fixed wage portion a such that $E[U^A] = 0$, or $a = -\frac{1}{8}b^2 \left((l + h)^2 + 2(l^2 + h^2) \right)$. Then $\frac{\partial E[U^P]}{\partial b} = \frac{1}{4}(1 - b) \left((l + h)^2 + 2(l^2 + h^2) \right)$ and hence $b = 1$. Thus the principal sells the value of the firm to the agent. This results in $e_1 = \frac{1}{2}(l + h) = e_1^{FB}$ and $e_2 = v = e_2^{FB}$.

The resulting utility levels of the agent at t_2 are given by $U^A(v = l) = \frac{1}{2}(l^2 - h^2) < 0$ and $U^A(v = h) = \frac{1}{2}(h^2 - l^2) > 0$. ■

Proof of Proposition (3). The agent's incentive compatibility constraint to pursue a newly offered \underline{m}_2 is given by (t_2 ICC), and the principal optimally sets $\underline{m}_2 = E^P[e_1] + v$. As discussed there are four possible scenarios that the agent's choice e_1 leads to, (i) he sticks to \underline{m}_1 , (ii) he can always be induced to pursue \underline{m}_2 , (iii) he can only be induced to pursue \underline{m}_2 if $v = l$ or only if $v = h$.

In scenario (i) the principal need never offer an \underline{m}_2, b_2 . The agent sets

$e_2 = \underline{m}_1 - e_1$ in order to meet \underline{m}_1 in expectation at $t = 2$, and hence maximizes $E[U^A] = b_1 - \frac{1}{2}e_1^2 - \frac{1}{2}(\underline{m}_1 - e_1)^2$. This gives first order condition $\frac{\partial E[U^A]}{\partial e_1} = \underline{m} - 2e_1 = 0 \Rightarrow e_1 = \frac{1}{2}\underline{m}_1$, and hence the participation constraint $E[U^A] \geq 0$ results in $b_1 \geq \frac{1}{4}\underline{m}_1^2$. The principal sets this constraint binding and hence maximizes $E[U^P] = \frac{1}{2}(l + h)\underline{m}_1 - \frac{1}{4}\underline{m}_1^2$, resulting in first order condition $\frac{\partial E[U^P]}{\partial \underline{m}_1} = \frac{1}{2}(l + h) - \frac{1}{2}\underline{m}_1 = 0 \Rightarrow \underline{m}_1 = (l + h)$. This results in first-best first period effort but not first-best second period effort. It is however not incentive compatible for the principal to not offer another \underline{m}_2, b_2 that is incentive compatible for the agent to accept, as will be shown below when we consider scenario (ii).

In scenario (ii) the principal optimally offers $\underline{m}_2 = E^P[e_1] + v$, with b_2 given by (t_2 ICC) set binding. At $t = 1$ the agent anticipates exerting $e_{2,v} = \underline{m}_2 - e_1$, such that he maximizes:

$$E[U_1^A] = \frac{1}{2} \left(b_{2,h} - \frac{1}{2} (h + E^A[E^P[e_1]] - e_1)^2 \right) + \frac{1}{2} \left(b_{2,l} - \frac{1}{2} (l + E^A[E^P[e_1]] - e_1)^2 \right) + b_1 - \frac{1}{2}e_1^2$$

where $b_{2,v} = \frac{1}{2} (m_2 - m_1) (m_1 + m_2 - 2e_1)$ or

$$b_{2,v} = \frac{1}{2} \left(-m_1^2 + 2m_1 E^A[E^P[e_1]] + v^2 - E^A[E^P[e_1]]^2 \right).$$

Since the agent does not influence the principal's expectation of e_1 this results in $\frac{\partial E[U_1^A]}{\partial e_1} = \frac{1}{2} (h + E^A[E^P[e_1]] - e_1) + \frac{1}{2} (l + E^A[E^P[e_1]] - e_1) - e_1 = 0$ or $e_1 = \frac{1}{4}(l + h) + \frac{1}{2}E^A[E^P[e_1]]$. In equilibrium the principal has the correct belief regarding $E^P[e_1]$, that is $E^P[e_1]^* = e_1$, and the agent has the correct belief regarding $E^P[e_1]$ such that $e_1 = \frac{1}{4}(l + h) + \frac{1}{2}e_1$ or $e_1 = \frac{1}{2}(l + h)$ resulting in the first-best outcome. Note also that this implies that e_2 is also equal to the first-best: $e_{2,v} = v$.

This could be taken to imply that no initial contract is required. However at $t = 2$ it is in the principal's interest not to reward the agent for efforts already provided at $t = 1$, effort that is sunk to the agent at $t = 2$. An initial contract

Appendix

\underline{m}_1 must be offered to prevent the principal from extorting the agent at $t = 2$, as the agent can anticipate being given no rent above the $t = 1$ contract. At $t = 1$ the principal should set up the initial contract, and the agent works such that he can meet that initial target, anticipating that he will not earn a rent due to the second period amended contract. Thus the agent at $t = 1$ maximizes $E[U_1^A] = b_1 - \frac{1}{2}e_1^2 - \frac{1}{2}(\underline{m}_1 - e_1)^2$, f.o.c. $\frac{\partial E[U_1^A]}{\partial e_1} = \underline{m}_1 - 2e_1 = 0$, or $e_1 = \frac{1}{2}\underline{m}_1$. This results in participation constraint $b_1 \geq \frac{1}{4}\underline{m}_1^2$. Setting that binding and substituting in the principal's first period maximization problem yields $\underline{m}_1 = l+h$.

Note that this is the same initial target as under scenario (i) and that it is incentive compatible for the principal to adjust the initial contract at $t = 2$. Furthermore if the agent exerts e_1 as described his prescribed second period efforts are incentive compatible. However it is not incentive compatible for the agent to exert effort as necessary for the contract to be revisited in both realizations of v . Consider that the principal expects the agent to pursue the new contract and leaves the agent no rent, hence total compensation equals cost of efforts. Let the principal expect the agent to exert $e_1 = e_1^P$, $e_{2,l} = e_{2,l}^P$ and $e_{2,h} = e_{2,h}^P$, where $e_{2,h}^P > e_1^P > e_{2,l}^P$. Consider following deviation from the agent: exert $e_1 = e_1^P - \Delta$ and $e_{2,l} = e_{2,l}^P + \Delta$, but not pursue \underline{m}_2 if $v = h$ such that $e_{2,h} = e_1$. Total compensation associated with the initial target must be $w^A(\underline{m}_1) = (e_1^P)^2$, and $w^A(\underline{m}_{2,l}) = \frac{1}{2}(e_1^P)^2 + \frac{1}{2}(e_{2,l}^P)^2$, expected utility from this deviation for the agent equals:

$$\begin{aligned} E[U^A] &= \frac{1}{2} \left(\frac{1}{2}(e_1^P)^2 + \frac{1}{2}(e_{2,l}^P)^2 - \frac{1}{2}(e_{2,l}^P + \Delta)^2 \right) \\ &\quad + \frac{1}{2} \left((e_1^P)^2 - \frac{1}{2}(e_1^P + \Delta)^2 \right) - \frac{1}{2}(e_1^P - \Delta)^2 \\ E[U^A] &= \frac{1}{2}\Delta(e_1^P - e_{2,l}^P - 2\Delta) \end{aligned}$$

Note that since $e_{2,l}^P < e_1^P$ and $\Delta > 0$ this yields $E[U^A] > 0$ if:

$$\frac{e_1^P - e_{2,l}^P}{2} > \Delta$$

Hence as long as the principal expects the agent to always pursue \underline{m}_2 the agent can profitably deviate (under scenario (ii) the agent earned no rent) by lowering efforts in the first period and only accepting the proposed second period contract if $v = l$. It is easily demonstrated that the proposed deviation is incentive compatible such that scenario (ii) cannot form an equilibrium.

In scenario (iii) the agent will pursue m_2 if $v = l$ but not if $v = h$, and hence optimally exerts $e_{2,l} = m_{2,l} - e_1$ and $e_{2,h} = m_1 - e_1$. The principal sets $m_{2,l} = l + E^P[e_1]$ and $b_{2,l}$ according to a binding (t_2 ICC). The agent at $t = 1$ solves $\max_{e_1} E[U_A] = \frac{1}{2}(-\frac{1}{2}e_{2,m_1}^2) + \frac{1}{2}(b_{2,l} - \frac{1}{2}e_{2,l}^2) + b_1 - \frac{1}{2}e_1^2$, which after substitution results in the f.o.c. $\frac{\partial E[U_A]}{\partial e_1} = \frac{1}{2}(m_1 - e_1) + \frac{1}{2}(m_{2,l} - e_1) - e_1 = 0$, where we utilize that the agent does not influence the principal's beliefs. This results in $e_1 = \frac{1}{4}(m_1 + E^A[m_{2,l}])$. The agent and principal can deduce one another's optimal action and hence $E^P[e_1] = \frac{1}{4}(m_1 + E^A[m_{2,l}]) = \frac{1}{4}(m_1 + E^A[l + E^P[e_1]])$, in equilibrium the beliefs must be correct or $E^P[e_1] = E^A[e_1] = e_1^*$, and therefore $e_1^* = \frac{1}{3}(m_1 + l)$.

Substituting all derived values in $E[U^A]$ gives $E[U^A] = b_1 - \frac{1}{6}l(4l + m_1)$, which determines the agent's participation constraint.

The principal sets $b_1 = \frac{1}{6}l(4l + m_1)$ and maximizes her utility:

$$\max_{m_1} E[U^P] = \frac{1}{3}l^2 + \frac{1}{6}lm_1 - \frac{1}{6}m_1^2 + \frac{1}{2}hm_1$$

$\frac{dE[U^P]}{dm_1} = \frac{1}{6}l - \frac{1}{3}m_1 + \frac{1}{2}h = 0 \iff m_1 = \frac{1}{2}l + \frac{3}{2}h$. Such that compared to scenario (i) and (ii) the principal sets a higher bar for m_1 to compensate for the agent's incentive to slack off in the first period. This results in $m_{2,l} = \frac{1}{2}h + \frac{3}{2}l$, $e_1 = \frac{1}{2}(h + l) = e_1^{FB}$, and $e_{2,v} = v = e_2^{FB}$.

Since the contract induces the optimal efforts in the second period it is not incentive compatible for the principal to offer an amended contract if $v = h$, and the agent's incentive compatibility constraint to pursue m_2 if $v = l$ is met by definition.

Scenario (iv) is symmetric to scenario (iii). ■

Appendix

Proof of Proposition 4. Let the initial contract be $w_1^A = a_1 + b_1 m$ and the renegotiated contract in the second period given v $w_2^A = a_2 + b_2 m$. The game is solved through backward induction.

At $t = 2$ the agent decides between accepting the new contract or continuing under the old. Under the old contract his utility is $U^A(w_1) = a_1 + b_1(e_1 + e_2) - \frac{1}{2}e_1^2 - \frac{1}{2}e_2^2$, and the agent optimally exerts effort according to $\partial U^A(w_1)/\partial e_2 = b_1 - e_2 \Rightarrow e_2 = b_1$. This results in $U^A(w_1) = a_1 + b_1(e_1 + b_1) - \frac{1}{2}e_1^2 - \frac{1}{2}b_1^2$.

Under the new contract his expected utility is $U^A(w_2) = a_2 + b_2(e_1 + e_2) - \frac{1}{2}e_1^2 - \frac{1}{2}e_2^2$, and $\partial U^A(w_2)/\partial e_2 = b_2 - e_2 \Rightarrow e_2 = b_2$ resulting in $U^A(w_2) = a_2 + b_2(e_1 + b_2) - \frac{1}{2}e_1^2 - \frac{1}{2}b_2^2$.

Thus it is best for the agent to accept the new contract if $U^A(w_2) \geq U^A(w_1)$ or:

$$a_2 - a_1 + (b_2 - b_1)e_1 \geq \frac{1}{2}(b_1^2 - b_2^2)$$

Hence, the income difference must be greater than or equal to the associated effort cost difference.

At $t = 2$ the principal when renegotiating maximizes utility under the condition that the $e_2 = b_2$ and that a_2 is such that $U^A(w_2) = U^A(w_1)$ since she has no incentive to give the agent a rent. Her utility is thus given by:

$$E[U_P] = v(E^P[e_1] + b_2) - b_2(E^P[e_1] + b_2) - a_1 - (b_1 - b_2)E^P[e_1] - \frac{1}{2}(b_1^2 - b_2^2)$$

Resulting in first order condition $\partial U_P/\partial b_2 = v - E^P[e_1] - 2b_2 + E^P[e_1] + b_2 = 0$ such that $b_2 = v$, that is the agent is paid the marginal product value. This implies that it is always in the principal's interest to renegotiate at $t = 2$ to realign the agent's incentives with the value of effort.

Note that the principal's belief regarding e_1 must be correct (or not off in the wrong direction) in order for the agent to be willing to accept the renegotiated contract. Thus the agent's first period effort choice and the principal's belief regarding that choice determine the equilibrium outcome. This can result in

the contract always being renegotiated or only being renegotiated in one of the realizations of v .

First consider when the contract is always renegotiated, in which case the agent's incentives are ultimately driven by $b_2 = v$. Anticipating that he will be held to his expected utility following from w_1 since the principal sets a_2^v such that the agent earns no rents. Hence the agent simply maximizes $U^A = a_1 + b_1(e_1 + b_1) - \frac{1}{2}e_1^2 - \frac{1}{2}b_1^2$, resulting in first order condition $dU^A/de_1 = b_1 - e_1 = 0$, or $e_1 = b_1$.

At $t = 1$ the principal must ensure that it is incentive compatible for the agent to accept the contract or $E[U^A] = a_1 + b_1^2 \geq 0$. Furthermore she expects to renegotiate the initial contract and offer $b_2 = v$ and a_2 such that $U^A(w_2) = U^A(w_1)$ in the second period. This implies that she essentially gives what value the agent creates to the agent but creams off a_2 . The principal's therefore maximizes:

$$E[U^P] = -\frac{1}{4}h^2 - \frac{1}{2}hb_1 - \frac{1}{4}l^2 - \frac{1}{2}lb_1 + \frac{1}{2}b_1^2$$

The first order condition $\partial E[U^P]/\partial b_1 = -\frac{1}{2}(h + l) + b_1 = 0$ results in $b_1 = \frac{1}{2}(l + h)$. This results best effort levels in the first and second period, and $E[U^A] = 0$ since the agent is always just reimbursed for his effort costs, while the principal's expected utility is $E[U^P] = \frac{3}{8}h^2 + \frac{1}{4}hl + \frac{3}{8}l^2$.

At $t = 2$ it is incentive compatible to accept the proposed w_2 given the first period effort derived above. However if the agent anticipates on sticking to the original contract w_1 in one realization of v he can profitably deviate. In scenario (ii) the principal anticipates $e_1 = b_1$, and $e_2 = b_2$, and offers contracts leaving no rents based on that expectation, hence the fixed wage associated with the initial contract equals $a_1 = -b_1^2$ and that associated with the second contract $a_{2,l} = \frac{1}{2}b_1^2 - b_1b_{2,l} - \frac{1}{2}b_{2,l}^2$. Consider the following strategy of the agent: exert $e_1 = b_1 - \Delta$ and do not renegotiate if $v = h$. The agent's expected utility in this

Appendix

scenario is:

$$\begin{aligned}
 E[U^A] &= \frac{1}{2} \left(b_1 (2b_1 - \Delta) - b_1^2 - \frac{1}{2} b_1^2 \right) \\
 &\quad + \frac{1}{2} \left(b_2 (b_1 - \Delta + b_2) + \frac{1}{2} b_1^2 - b_1 b_{2,l} - \frac{1}{2} b_{2,l}^2 - \frac{1}{2} b_{2,l}^2 \right) \\
 &\quad - \frac{1}{2} (b_1 - \Delta)^2 \\
 E[U^A] &= \frac{1}{2} \Delta (b_1 - b_{2,l} - \Delta)
 \end{aligned}$$

Note that $E[U^A] > 0$ if $b_1 - b_{2,l} > \Delta$, and since $b_{2,l} < b_1$ in the scenario considered it is always possible to find a deviation Δ that yields the agent a rent. Since the agent can profitably deviate scenario (ii) is not an equilibrium.

Now consider the scenario in which the agent and principal anticipate on the contract not being renegotiated when one of the realizations of v occurs. As an example we take the scenario in which the principal does not propose a new contract if $v = h$, the other case is symmetric. The agent and principal's optimal actions at $t = 2$ are not changed and the agent optimally sets $e_1 = \frac{1}{2} (l + b_1)$ as determined above. Note that both the agent and the principal can deduce the optimal e_1 , hence we should have $E^A[E^P[e_1]] = E^P[e_1] = e_1^* = \frac{1}{2} (b_1 + l)$. Substituting the outcome variables into $E[U^A]$ gives the participation constraint $a_1 \geq -\frac{1}{8}l^2 - \frac{1}{4}lb_1 + \frac{1}{4}l - \frac{7}{8}b_1^2$, which the principal optimally sets binding. Thus the principal's expected utility at $t = 1$ is:

$$\begin{aligned}
 E[U^P] &= \frac{1}{2} (-a_2^l) + \frac{1}{2} ((h - b_1) (e_1 + e_2^h) - a_1) \\
 E[U^P] &= \frac{3}{4}hb_1 - \frac{1}{4}l - \frac{3}{8}b_1^2 + \frac{1}{4}hl + \frac{5}{8}l^2
 \end{aligned}$$

Maximization of which yields $\frac{dE[U^P]}{db_1} = \frac{3}{4}h - \frac{3}{4}b_1 = 0$ or $b_1 = h$. Substitution shows that $e_2^h = h$, $e_2^l = l$ and $e_1 = \frac{1}{2} (h + l)$ such that the first-best outcome is achieved. Note that $a_2^l = \frac{1}{8}h^2 - \frac{1}{4}hl - \frac{9}{8}l^2 + \frac{1}{4}l$ is set such that the ICC if $v = l$ holds, and no other contract needs to be offered if $v = h$. Note that the fixed wage $a_1 = -\frac{1}{8}l^2 - \frac{1}{4}lh + \frac{1}{4}l - \frac{7}{8}h^2$ protects the agent from any possible extortion.

■

Chapter 2

Biased Supervision

*Joint with Josse Delfgaauw*¹

2.1 Introduction

In many organizations, middle managers' assessment of employees' performance is an important determinant of bonus pay and career prospects.² If verifiable performance measures are imperfect, subjective performance evaluation may provide a more accurate assessment of employees' performance, thereby providing better incentives for employees to perform well. On the other hand, by its very nature, subjective performance evaluation can be manipulated, weakening the link between actual and reported performance. Furthermore, managers' role in determining employees' bonus pay and promotion opportunities gives them (more) power over their subordinates. Earlier work has shown that subjective performance pay based on middle managers' evaluations can be prone to favoritism (Prendergast and Topel 1996, Bol 2011, Dur and Tichem 2015), collusion (Tirole

¹This chapter has been published in *Journal of Economic Behavior & Organization* (2016), 130, 107-125.

²For instance, Eccles and Crane (1988), Gibbs (1995), and Bol (2011) document the use of subjective performance evaluation in (financial) service firms, Breuer et al. (2013) in a large call-center, Gibbs et al. (2004) in car-dealerships, Woods (2012) in an internal audit firm, and Medoff and Abraham (1980) in manufacturing firms.

1986, Thiele 2013), extortion (Laffont 1990, Vafai 2002, 2010), and a lack of incentives or ability to monitor on the side of the manager (Gibbs et al. 2004, Bol 2011, Kamphorst and Swank 2018).

In this paper, we show that a supervisor can use the discretion inherent in subjective performance evaluation to pull agents towards tasks that benefit the supervisor more than the organization. We develop a principal-supervisor-agent model, where the agent exerts effort on multiple tasks. Efforts are not observable to the principal. The supervisor observes the agent’s efforts with a probability that is increasing in the supervisor’s ability. The supervisor provides a report on the agent’s efforts to the principal, which can be used in determining the agent’s (incentive) pay. Crucially, we assume that the supervisor has an intrinsic preference for particular tasks exerted by the agent. This makes that she overemphasizes these tasks when providing directions to the agent. Anticipating that not living up to the supervisor’s expectations results in a bad evaluation, the agent works towards the supervisor’s goals. When the supervisor’s preferences are not aligned with the principal’s goals, this hurts the principal. As a consequence, akin to the standard multitasking model (Holmström and Milgrom 1991, Baker 1992, 2002), the principal optimally sets weaker subjective performance pay when the supervisor’s preferences are less aligned, as well as when the supervisor has lower ability.³

This changes when the principal has access to a verifiable, but possibly incongruent, performance measure. To structure ideas, consider one salesman of a local store owned by an electronics retail chain. The store’s manager is an active member of the local community, so that she cares a lot about her store’s reputation for providing good service. The salesman contributes to long-run store performance through sales effort and service effort. The latter does not contribute directly to short-run sales, but increases the reputation of the local store,

³Supervisors can also use their power to affect (the behaviour of) employees in ways that are not directly linked to employees’ tasks at work, e.g. by engaging in bullying, extortion, (sexual) harassment, etc. Our interest lies with supervisors’ incentives to provide misaligned directions regarding employees’ efforts at work.

which has long-run benefits to the retail chain. The store manager typically observes efforts, but the chain's headquarters only observes sales. If headquarters uses the salesman's sales figures to provide incentive pay, he will focus his efforts disproportionately on sales at the expense of service, leading to a sub-optimal outcome. Alternatively, headquarters could relate the salesman's pay to his performance evaluation as provided by the store manager. However, in evaluating performance, the store manager will put too much emphasis on service provision, thereby inducing the salesman to exert suboptimally low sales effort (from the perspective of headquarters). Combining verifiable sales figures with subjective performance evaluation in the salesman's bonus plan may bring several advantages. First, headquarters can use sales targets, which constrains the store manager in emphasizing service at the expense of sales. Second, the inclusion of subjective performance evaluation allows the store manager to pull the salesman away from the disproportionate focus on sales induced by the verifiable sales measure. Third, the sales figures may give the store manager additional information on the salesman's efforts, which allows for better monitoring.⁴

We show that combining verifiable performance measurement and subjective performance evaluation mitigates the distortion that arises when using either type of performance measurement exclusively. This echoes findings in the multitasking literature on combining multiple incongruent performance measures (Feltham and Xie 1994, Datar et al. 2001, Budde 2007). This literature has shown that full congruence can be achieved if the number of verifiable measures meets or exceeds the number of tasks (although full congruence is not optimal if the measures differ in noisiness and agents are risk-averse). Even when all measures are biased towards the same task, congruence is possible by placing a negative weight on the most biased measure. In contrast, we show that this does not hold when some measures are subjectively determined, because placing a negative weight on the subjective evaluation is ineffective. The supervisor uses any discretionary power

⁴For an example of an incentive plan that combines verifiable performance measures and subjective performance evaluation in a retail setting, see Bouwens and Kroos (2011).

to put more emphasis on the tasks she considers undervalued in the objective performance measures. If a good evaluation has a positive (negative) effect on the agent's compensation, the supervisor threatens to provide a bad (good) evaluation unless the agent follows her directions. Hence, congruence is not feasible when the supervisor is more biased than the verifiable performance measure.

The principal prefers the verifiable performance measure and the supervisor to have opposite biases compared to his relative valuation of tasks. We show that this allows the principal to implement non-distorted efforts by offering a bonus conditional on achieving both a performance target and a favourable subjective evaluation, unless either the performance measure or the supervisor is relatively ineffective. We model ineffectiveness of the performance measure as the probability that the agent can ex post costlessly manipulate measured performance. If this probability is too high, the supervisor would ignore the principal's performance target and induce her most preferred effort allocation, requiring the agent to manipulate measured performance when possible. If the supervisor's ability is too low, the agent would ignore her instructions and meet the performance target at lowest effort cost by working purely towards measured performance. To prevent these outcomes, the principal optimally adjusts the performance target to allow for some bias in implemented efforts and reduces the strength of the agent's incentive pay.

The key novelty of our model is that the supervisor has intrinsic preferences over her subordinates' tasks, which may differ from the principal's relative valuation of these tasks. Such preferences could be driven by private benefits, by career concerns, or by professional norms. The supervisor may overemphasize providing inputs into her own work, thereby reducing her own workload. In a multi-unit organization, the supervisor could overemphasize tasks that benefit the supervisor's unit at the expense of activities that benefit other units. Alternatively, the supervisor may intrinsically consider particular tasks more important. Akerlof and Kranton (2005) suggest that internalization of norms is an important element of professional training (of e.g. physicians, scientists, and teachers). Professional

Introduction

norms may guide what is considered to be worthy of doing, and supervisors may impose these norms on their subordinates. Relatedly, Prendergast (2007) considers bureaucrats who care more or less for clients' well-being than the principal, possibly leading to over- or underprovision of services to clients.⁵

In our setup, a biased supervisor uses her discretionary power to direct employees towards activities that benefit her. Middle managers questioned by Guth and MacMillan (1986) stated that they sometimes make decisions that are not aligned with corporate strategy and goals, in order to protect their self-interest. Burgelman (1994) describes events at Intel in the 1980s, where middle managers made decisions on the allocation of R&D and manufacturing capacity that went against corporate strategy, eventually forcing senior management to change strategy.⁶ Our analysis shows that misaligned middle management may, but need not be detrimental for firm performance, depending on the available verifiable performance measures.

Most earlier work on combining subjective and objective performance measures considers subjective evaluation by the principal (Baker et al. 1994, Schmidt and Schnitzer 1995, Pearce and Stacchetti 1998, Budde 2007). Following Bull (1987), the emphasis lies on self-enforcing relational contracts, where the size of the subjectively determined bonus is restricted by the principal's incentive to give low evaluations despite good performance in order to save on bonus payments.⁷ Thiele (2013) shows that in this setting delegation of subjective performance evaluation to a supervisor also entails low-powered incentives, in order to prevent

⁵The supervisor's bias in preferences over tasks differs from interpersonal preferences towards employees, such as altruism, spite or favouritism, as studied by e.g. Prendergast and Topel (1996), Giebe and Gürtler (2012), and Dur and Tichem (2015). These papers show that interpersonal preferences mute the incentive effect of subjective performance pay by weakening the link between agent's effort and pay.

⁶Rotemberg and Saloner (2000) provide a rationale for giving middle managers some discretion to go against corporate strategy. They argue that combining a visionary CEO, who champions some units or product lines over others, with more neutral middle management provides strong incentives to come up with investment projects for employees in the favoured units, while if they nonetheless fail to find suitable projects, middle managers can allocate funds towards projects in other units.

⁷Suvorov and Van de Ven (2009) and Zabochnik (2014) show that the incentive to underreport is considerably smaller when the principal also uses performance evaluation to provide the agent with information about his ability.

collusion. Following Tirole (1986), collusion is also the main issue studied in static three-tier hierarchy models; for overviews see Laffont and Rochet (1997) and Mookherjee (2013). In this paper, we assume away the problem of collusion in our static model by assuming that side-contracts are not enforceable. Instead, we focus on the effects of supervisor bias on the interaction between subjective and objective performance evaluation.

Most related to our work are Laffont (1990) and Vafai (2002, 2010), who study abuse of authority by the supervisor. In a setting with two agents and hard information on total output, Laffont (1990) shows that a supervisor can extort favors or side-payments from her subordinates by threatening to manipulate individual performance information. In response, the principal optimally reduces the weight on individual performance as reported by the supervisor in the agent's incentive pay, relative to the weight on total output. In Vafai (2002, 2010), the supervisor may receive hard information about the agent's performance and can decide to conceal this information from the principal. Vafai (2002) shows that the supervisor can collude with the agent when performance is bad, and can demand a bribe from the agent when performance is good. Preventing this abuse of authority requires that the agent's pay is not higher after a report showing good performance compared to a report without performance information. This also prevents collusion, but leaves a rent to the agent. Vafai (2010) extends the analysis to side-payments from the principal to the supervisor. In our setting, abuse of authority arises from the supervisor's soft information on agent's efforts and materializes in a distorted effort allocation. We analyze how the principal can use (imperfect) verifiable performance measures to constrain the supervisor's abuse of authority.⁸

We use our results to contribute to the debate on the relative merit of specialists and generalists in managerial positions. In the typical trade-off, specialists have higher ability on a particular task, but generalists have broader skills. Gar-

⁸In a setting where supervisors can conceal hard information, Kofman and Lawarree (1996) show that assigning a second supervisor makes it possible to prevent collusion.

icano (2000) argues that more able agents should be assigned to higher positions in the hierarchy in a one-dimensional task model, so that lower-level generalists can screen for tasks that can only be properly conducted by specialists. In contrast, Ferreira and Sah (2012) consider communication between layers and argue that generalists should be higher in the hierarchy to facilitate information transmission with lower-level units consisting of (different) specialists. Prasad (2009) argues that in a setting where tasks are complements, generalists are more likely to work on multiple tasks than specialists. He finds supporting evidence among non-academic researchers with doctoral degrees in the US, for whom the probability of getting management tasks is decreasing in past research success.

We argue that while specialists may have better monitoring ability, they are also more likely to have biased preferences (for instance arising from professional norms).⁹ We show that in the absence of verifiable performance measures, this gives a trade-off between the strength of subjective performance incentives and the distortion induced in the agent's efforts. Hence, a generalist manager is preferred when a specialist's objectives are too misaligned with the principal's, while a specialist is preferred when the generalist is too ineffective at supervision. The availability of a verifiable performance measure decreases both the cost of supervisor bias and the benefit of better monitoring. We find that the first effect typically outweighs the second effect, so that better verifiable performance measures increase the relative attractiveness of specialist supervisors.

The paper is organized as follows. The setup of the model is given in Section 2, after which we analyze benchmark cases in Section 3 and the full model in Section 4. We apply the results to the choice between specialists and generalists for managerial positions in Section 5. Section 6 concludes.

⁹This corresponds to Li (2013), who studies the decisions by reviewers of grant proposals at the US National Institute of Health, and finds that reviewers are both biased towards projects in their own area as well as better able to infer the quality of proposals in their own area.

2.2 The model

We consider a principal-supervisor-agent model in which all players are risk neutral. The principal (P) employs one agent (A) and one supervisor (S). The outside option utility of both the agent and the supervisor is zero, and they are both protected by limited liability such that $w_A \geq 0$ and $w_S \geq 0$, where w_A (w_S) is the total wage payment to the agent (supervisor).¹⁰

The agent works on two tasks $i \in \{1, 2\}$. The principal values the two tasks equally, and his utility is given by

$$U_P = e_1 + e_2 - w_A - w_S \quad (2.1)$$

where e_i is the agent's effort in task i . The agent's utility is given by:

$$U_A = w_A - \frac{1}{2} (e_1)^2 - \frac{1}{2} (e_2)^2 \quad (2.2)$$

The principal cannot observe the agent's efforts. However, there is a verifiable but imperfect performance measure of the agent's efforts. The principal can use this verifiable performance measure to provide the agent with performance-related pay, as will be discussed below. However, the measure is imperfect in two ways, akin to the dimensions of distortion and noise in e.g. Feltham and Xie (1994) and Baker (2002). First, the performance measure is biased towards one of the tasks. The level of measured performance m is given by

$$m(e_1, e_2) = \varphi e_1 + (1 - \varphi) e_2 \quad (2.3)$$

with $\varphi \in [0, 1]$. Hence, if $\varphi \neq \frac{1}{2}$, the relative importance of the two tasks in determining measured performance differs from the relative valuation of the tasks by the principal. Second, the objective performance measure is not perfectly reliable. We model this as follows. After choosing efforts, with probability $1 - q$

¹⁰Limited liability also implies that the principal cannot sell the firm to the agent.

The model

the agent can costlessly manipulate measured performance into showing any level of m as preferred by the agent. The principal cannot detect this type of gaming. As discussed below, the use of a less reliable performance measure (lower q) implies higher rents for the agent.¹¹

The only role of the supervisor is to monitor the agent. The supervisor observes the agent's actual effort choice with probability p , reflecting the supervisor's effectiveness in monitoring the agent. The supervisor's information is soft and cannot be made verifiable, but she can make a verifiable report r regarding her assessment of the agent's performance. It follows that the supervisor's report is cheap talk: she can provide any report independent of the agent's actual efforts.¹² When making her report, the supervisor has access to the signal provided by the verifiable performance measure.¹³ She does not observe directly whether the agent manipulated measured performance (but can infer this when observing effort). The supervisor's report can provide a basis for (subjective) performance pay, as the principal can make the agent's pay dependent on the report r .

Crucially, we assume that the supervisor cares about the tasks performed by the agent, possibly attaching different (relative) weights to the tasks as compared to the principal. As discussed in the Introduction, these preferences may stem from career concerns, professional norms, or intrinsic care for the tasks' output. The supervisor's utility is given by:

$$U_S = w_S + \eta e_1 + (1 - \eta) e_2 \quad (2.4)$$

¹¹The assumption of (un)reliability provides a tractable way of introducing a second dimension of performance measure imperfection into our model (on top of the bias). This allows us to compare our results with earlier work that studies the trade-off between distortion and noise in performance measures. There are several alternative ways of modeling unreliability, including a setting where the agent can always costlessly manipulate measured performance while the principal detects gaming with probability q , and a setting where the objective performance measure fails to provide any measured performance with probability $1 - q$.

¹²This differs from Tirole (1986) and Vafai (2010), where the supervisor reports either the true (outcome of the agent's) efforts or reports that she received no information.

¹³Supervisors typically have access to the verifiable performance information when evaluating employees. For instance, Bol (2011) shows that the form supervisors had to fill out contained both verifiable and subjective items, where the items based on verifiable measures came with strict guidelines on how to translate performance into rating. The results are qualitatively similar when assuming that the supervisor does not observe m before determining r .

where $\eta \in [0, 1]$. Hence, the supervisor has biased preferences relative to the principal's valuation of tasks whenever $\eta \neq \frac{1}{2}$.¹⁴ Note that the supervisor does not incur monitoring cost.

Our assumption of limited liability ($w_S \geq 0$) implies that the principal cannot acquire the rents from intrinsic utility obtained by the supervisor. We make this assumption to ensure that the optimal contract as designed by the principal is aimed at optimizing the agent's incentives, rather than at increasing the supervisor's intrinsic utility. Furthermore, we assume that the principal does not provide incentive pay to the supervisor. Incentive pay would have to be based on the verifiable measure m , which would imply that the supervisor's relative preferences over tasks would be drawn closer to the relative weights on tasks in m . We argue in Section 2.4 that this is typically not in the principal's interest. Hence, the principal offers a fixed wage $w_S = 0$, which the supervisor accepts.

The discussion above implies that the agent's wage can depend on both measured performance m and the supervisor's report r . For expositional reasons we explicitly distinguish between purely objective performance pay $b(m)$ and subjective performance pay $c(m, r)$, such that the agent's wage equals $w_A(m, r) = b(m) + c(m, r)$.¹⁵

Provided that the agent's wage depends on the supervisor's evaluation, the supervisor can make demands to the agent. We assume that side-contracts are not verifiable (neither between the supervisor and the agent nor between the supervisor and the principal).¹⁶ Instead, the supervisor and the agent engage in an implicit agreement. As the supervisor does not bear the cost of the agent's bonus pay, she is ex post indifferent between providing different subjective evaluations.

¹⁴In our analysis, only the relative weights of the supervisor's preferences over tasks matters, not the absolute level. Hence, we could multiply the last two terms of (2.4) with the same parameter without affecting results.

¹⁵On top of this, the principal may offer fixed wage a to the agent, so that $w_A = a + b(m) + c(m, r)$. However, our assumption of limited liability ($w_A \geq 0$) makes that the principal always sets $a = 0$. It also implies that the agent's participation constraint is always fulfilled.

¹⁶This makes collusion non-sustainable. The bonus is paid to the agent after the supervisor has reported to the principal, implying that the agent has no incentive to transfer part of the bonus to the supervisor. Collusion is studied by e.g. Tirole (1986) and Thiele (2013).

The model

We assume that the supervisor adheres to the implicit agreement.¹⁷ Given that the supervisor's evaluation is cheap talk, her demands have most influence when she promises to provide evaluations such that her report has a maximal effect on the agent's wage. Without loss of generality, we assume that the supervisor demands effort levels \underline{e}_1 and \underline{e}_2 , and promises to provide in return the report that yields the highest wage (given measured performance m). If the supervisor learns that the agent did not adhere to the demand (either from own observation or when the verifiable performance measure shows that performance is incompatible with the supervisor's demands), she provides the report that yields the lowest possible wage. We assume that if the supervisor is unsure about the agent's efforts, which happens when she does not observe effort and measured performance m is in line with demanded performance (which could be the result of manipulation) she provides the report that yields the highest subjectively determined bonus.¹⁸ Effectively, this implies that the supervisor will provide one of two reports. Hence, we can reduce the set of possible reports to two, $r \in \{r_G, r_B\}$. Note that the difference in bonus pay following a good subjective evaluation r_G and a bad subjective evaluation r_B may depend on measured performance m .

The timing of the game is as follows:

1. The principal offers the agent a contract, determining $w_A(m, r)$.
2. The agent accepts or rejects the contract. If the agent rejects, all players receive their outside option payoff.
3. The supervisor demands effort levels $\{\underline{e}_1, \underline{e}_2\}$ from the agent.
4. The agent chooses effort, which is observed by the supervisor with probability p .

¹⁷In a repeated game, adhering to the implicit agreement would be strictly preferred by the supervisor, to maintain credibility.

¹⁸This assumption implies that ineffective supervision (low p) yields rents to the agent. Alternatively, the supervisor might send a report that tells the principal that she is unsure. It is easily shown that, compared to the case studied in the paper, the principal would reduce the agent's rents by offering $w_A(m) = 0$ after this report. We make this assumption on the supervisor's reporting strategy in order to be able to study the effects of supervisor ability (p) on effort. If the supervisor would report to be unsure, equilibrium effort would not be a function of p . All other results would be qualitatively similar.

5. With probability $1 - q$, the agent can manipulate measured performance m .
6. The supervisor observes m and sends report $r \in \{r_G, r_B\}$.
7. Payoffs are realized.

We use backward induction to solve for a sub-game perfect Nash equilibrium.

2.3 Benchmarks

2.3.1 Complete information

Suppose the principal can contract on effort directly. Neither the supervisor nor the performance measure have any use in this case. The principal demands the effort levels that maximize his utility (2.1) subject to the agent's participation constraint $U_A \geq 0$, where U_A is given by (2.2). Ignoring the supervisor's intrinsic utility, this gives first-best levels of effort $e_1 = e_2 = 1$. The participation constraint of the agent is satisfied by setting $w_A = 1$ if and only if $e_1 = e_2 = 1$ and $w_A = 0$ otherwise. This results in $U_P = 1$ and $U_A = 0$. Hence, in the absence of moral hazard problems, the principal optimally induces the agent to balance effort levels across tasks.

2.3.2 Pure objective performance pay

In the absence of subjective performance evaluation, the model is a standard multitasking model. Without loss of generality, we assume that the principal offers the agent a fixed bonus b if measured performance (2.3) is above a specified target, $m \geq \underline{m}$, and no bonus if $m < \underline{m}$. Using backward induction, if the agent gets the opportunity to manipulate measured performance, he secures his bonus by setting $m = \underline{m}$ if his true performance is below \underline{m} .

In choosing effort, the agent derives no benefits from outperforming the principal's target \underline{m} . Hence, conditional on meeting the principal's target, the agent's

Benchmarks

optimal effort levels maximize (2.2) subject to $m(e_1, e_2) = \underline{m}$. This gives

$$e_1 = \frac{\varphi}{\varphi^2 + (1 - \varphi)^2} \underline{m} \quad (2.5)$$

$$e_2 = \frac{1 - \varphi}{\varphi^2 + (1 - \varphi)^2} \underline{m} \quad (2.6)$$

It follows that $e_1 = e_2$ only when $\varphi = \frac{1}{2}$. If $\varphi \neq \frac{1}{2}$, the agent provides more effort on the task that impacts measured performance m most. Given that the agent satisfies $m = \underline{m}$, he optimally chooses an effort combination such that

$$\frac{e_2}{e_1} = \frac{1 - \varphi}{\varphi} \quad (\text{PMR})$$

which we refer to as the Performance Measure's Ratio (PMR).

If the agent decides not to meet the performance target \underline{m} , he only receives bonus b if he can manipulate measured performance, which happens with probability $1 - q$. As this probability is independent of effort, optimal effort is zero. Hence, using (2.2), the agent chooses to meet the performance target if

$$b - \frac{1}{2} \frac{1}{\varphi^2 + (1 - \varphi)^2} \underline{m}^2 \geq (1 - q)b \quad (2.7)$$

which shows that the agent's rents $(1 - q)b$ are decreasing in the reliability of the performance measure.

The principal chooses b and \underline{m} to maximize his utility (2.1) subject to the agent's incentive compatibility constraint (2.7), which yields the following solution:

$$b = \frac{1}{2} \frac{q}{\varphi^2 + (1 - \varphi)^2} \quad (2.8)$$

$$\underline{m} = q \quad (2.9)$$

Regardless of the bias of the performance measure the principal demands $\underline{m} = q$. The principal demands lower total effort when the performance measure is more

biased.¹⁹ This is reflected in the bonus paid to the agent, which is higher when the performance measure is more aligned with the principal's objective. Equilibrium payoffs are given by

$$U_A = \frac{1}{2} \frac{q(1-q)}{\varphi^2 + (1-\varphi)^2} \quad (2.10)$$

$$U_P = \frac{1}{2} \frac{q}{\varphi^2 + (1-\varphi)^2} \quad (2.11)$$

The principal benefits from having a more effective and better aligned performance measure, as U_P increases in q and decreases in $|\frac{1}{2} - \varphi|$. Note that the agent's rent increases as the alignment of the measure and the principal's valuation of tasks improves. It is optimal for the principal to induce more effort the more aligned the measure, despite the higher rents to the agent. For a given φ , the sum of U_A and U_P is maximal when $q = 1$, reflecting that the principal sacrifices surplus to reduce the agent's rent when $q < 1$.

2.3.3 Pure subjective performance pay

Without objective performance measures, incentive pay is solely based on the supervisor's subjective report r . In the implicit agreement between the supervisor and the agent, the supervisor promises to provide a positive report to the principal if the agent performs at least effort levels \underline{e}_1 and \underline{e}_2 . The agent follows the supervisor's demands when a good report yields a sufficiently higher bonus than a bad report. Given limited liability, the optimal incentive scheme has wage zero after a bad report, $c(r_B) = 0$. Below, we denote $c(r_G) = c$.

The least costly way for the agent to satisfy the supervisor's demands is to provide exactly the demanded effort levels. If the agent decides to exert lower effort levels, he only receives the bonus when the supervisor does not observe the agent's efforts, which happens with probability $1 - p$. In this case, the best alternative is to provide no effort at all. Hence, the supervisor sets \underline{e}_1 and \underline{e}_2

¹⁹Using (2.5) and (2.6), it is easily seen that the sum of efforts $e_1 + e_2$ is decreasing in $|\frac{1}{2} - \varphi|$.

Benchmarks

to maximize her utility (2.4), subject to the agent's incentive compatibility constraint:

$$c - \frac{1}{2} \underline{e}_1^2 - \frac{1}{2} \underline{e}_2^2 \geq (1 - p) c \quad (2.12)$$

In order to induce the agent to exert effort, the agent must be given a rent equal to $(1 - p) c$. This follows from the fact that the supervisor does not always observe whether the agent performs as desired, so that shirking by the agent could pass undetected. The incentive compatibility constraint binds as the supervisor values additional effort. The supervisor's optimal effort demands are given by:

$$\underline{e}_1 = \eta \sqrt{\frac{2pc}{\eta^2 + (1 - \eta)^2}} \quad (2.13)$$

$$\underline{e}_2 = (1 - \eta) \sqrt{\frac{2pc}{\eta^2 + (1 - \eta)^2}} \quad (2.14)$$

It follows that \underline{e}_1 and \underline{e}_2 are increasing in bonus c and in the supervisor's effectiveness p . Both allow the supervisor to make stronger demands. Furthermore, the supervisor induces the agent to focus disproportionately on her preferred task, demanding effort levels such that:

$$\frac{\underline{e}_2}{\underline{e}_1} = \frac{1 - \eta}{\eta} \quad (\text{SR})$$

which we will refer to as the Supervisor's preferred effort Ratio (SR).

It follows from (2.13) and (2.14) that for given c and p , the principal's valuation of implemented efforts, $\underline{e}_1 + \underline{e}_2$, is higher when the supervisor's preferences are more aligned with the principal, i.e. when η is closer to $\frac{1}{2}$. Hence, a more aligned supervisor induces higher total effort for a given bonus, which benefits the principal.

The principal chooses c to maximize utility (2.1), taking into account the supervisor's effort demands (2.13) and (2.14). The optimal subjective bonus is given by:

$$c = \frac{1}{2} \frac{p}{\eta^2 + (1 - \eta)^2} \quad (2.15)$$

The bonus is increasing in p , as more effective supervisors leave fewer rents to the agent. The bonus also increases in the alignment between the supervisor and the principal, as more aligned supervisors use their discretionary power to demand higher total effort. Equilibrium effort levels are given by:

$$e_1 = \frac{\eta p}{\eta^2 + (1 - \eta)^2} \quad (2.16)$$

$$e_2 = \frac{(1 - \eta) p}{\eta^2 + (1 - \eta)^2} \quad (2.17)$$

Hence, if $p \neq 1$ or $\eta \neq \frac{1}{2}$, effort provision under subjective pay for performance is below first-best levels.

The agent's and principal's equilibrium payoffs are given by:²⁰

$$U_A = \frac{1}{2} \frac{p(1 - p)}{\eta^2 + (1 - \eta)^2} \quad (2.18)$$

$$U_P = \frac{1}{2} \frac{p}{\eta^2 + (1 - \eta)^2} \quad (2.19)$$

Note that the payoffs of the agent and the principal are identical to their payoffs when using pure objective pay for performance (see (2.10) and (2.11)), with $p = q$ and $\eta = \varphi$ (or $\eta = 1 - \varphi$). The principal responds to biased supervision by reducing subjective performance pay, as in case of incentive pay based on the incongruent verifiable performance measure. Hence, biased supervision is as harmful for the principal (and the agent) as an incongruent performance measure.²¹

²⁰The supervisor's utility equals $U_S = p$ and, hence, is independent of her preferences over tasks. The higher utility attained by a more biased supervisor for a given bonus c is exactly offset by the reduction in the bonus by the principal. The supervisor's limited liability constraint makes that her utility is irrelevant to the principal. Without limited liability, the principal would distort the bonus in order to increase the supervisor's intrinsic utility, which he would capture through a negative base wage.

²¹If the supervisor's bias is private information, such that the principal only knows the distribution of η but not the current supervisor's bias, a similar result obtains. From effort levels (2.16) and (2.17), it follows that given bonus c , the principal's payoff is decreasing and convex in supervisor bias $|\frac{1}{2} - \eta|$. Hence, for any distribution of bias $|\frac{1}{2} - \eta|$, the principal optimally sets a bonus below the optimal bonus in case of a supervisor with average bias. Hence, uncertainty about supervisor bias leads to (even) weaker subjective performance pay.

2.3.4 Pure subjective versus pure objective pay for performance

Comparing equilibrium effort levels under pure objective performance pay, as given by (2.5) and (2.6) with $\underline{m} = q$, and subjective pay for performance, as given by (2.16) and (2.17), shows that objective performance pay yields the same outcome as subjective performance pay if $p = q$ and $\eta = \varphi$ or $\eta = 1 - \varphi$. Then, both means of incentivizing the agent yield the same distortion in efforts and have the same quality of monitoring.

More generally, suppose that the principal must choose between implementing either objective or subjective performance pay. Comparing the principal's payoffs (2.11) and (2.19) shows that the principal is better off using discretionary performance pay when:

$$\frac{p}{q} > \frac{\eta^2 + (1 - \eta)^2}{\varphi^2 + (1 - \varphi)^2} \quad (2.20)$$

Even if the supervisor is more biased than the performance measure (e.g. when $\frac{1}{2} < \varphi < \eta$), subjective performance pay is still preferred when the supervisor is sufficiently more effective than the performance measure, and vice versa. As the right-hand side of (2.20) can only take values between $\frac{1}{2}$ and 2, it follows that subjective performance pay is always preferred to objective performance pay when $p > 2q$, whereas objective performance pay is always preferred if $q > 2p$. Figure 2.1 gives the preferred type of performance pay for given values of η , φ , p , and q .²²

²²It is also possible to interpret Figure 2.1 as a comparison between supervisors who differ in bias and effectiveness.

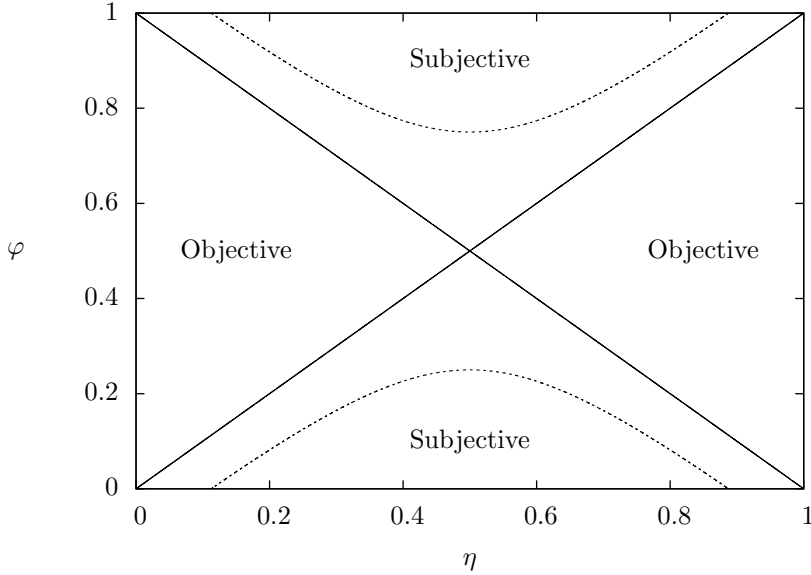


Figure 2.1: Preferred pure incentive mechanism if $p = q$ (solid) and $p = \frac{4}{5}q$ (dashed).

2.4 Subjective and objective performance pay combined

Now consider the case where both objective and subjective performance measures are available. This has several implications. First, the supervisor can pull the agent away from following the bias inherent in the objective performance measure. Intuitively, she will pull the agent closer to her own optimal effort ratio. Second, the principal may be able to use the objective performance measure to constrain the supervisor. Third, objectively measured performance gives the supervisor additional information on the agent's effort. This reduces the agent's opportunities to shirk and, hence, his rents.

We first establish several general features of the optimal contract.

Lemma 1 *The optimal contract is a forcing contract, where the agent receives compensation unless either objectively measured performance m differs from a*

pre-determined target \underline{m} or the supervisor reports bad performance r_B .

Proof. The proof is given in the Appendix. ■

To understand Lemma 1, consider first a supervisor with the same bias as the objective performance measure, $\eta = \varphi$. The use of objective performance pay leads to an outcome where $\frac{e_2}{e_1} = \frac{1-\varphi}{\varphi}$, as this effort ratio allows the agent to reach a given level of measured performance m at minimal effort cost (see (PMR)). Adding subjective performance pay does not affect this bias, as the supervisor also prefers to induce this effort ratio (given $\eta = \varphi$, see (SR)). There is still a benefit to the principal of conditioning the agent's pay on both objectively measured performance and subjective performance evaluation, as it yields the highest probability of detecting a shirking agent. The forcing contract minimizes the agent's rents. This benefit of combining objective and subjective performance evaluation carries over when the supervisor's preferences differ from the bias in the objective performance measure, i.e. when $\eta \neq \varphi$. Then, the supervisor uses any subjectively determined bonus to pull the agent towards her own preferred effort ratio. By conditioning the payout of the subjective bonus on objectively measured performance, the principal also constrains the supervisor. Denying the agent the bonus despite a positive report if measured performance does not meet a pre-determined target \underline{m} reduces the supervisor's power to demand efforts that fall short of the target.

A direct implication of Lemma 1 is that there is no clear separation between objective and subjective performance pay. Optimally, the agent's compensation depends on both objectively measured performance as well as the supervisor's subjectively determined report. For instance, the contract could have verifiable performance pay $b(m \neq \underline{m}) = 0$ and $b(\underline{m}) > 0$, with the 'disqualifier' that the bonus is forfeit after a bad subjective evaluation: $c(m, r_B) = -b(m)$ and $c(m, r_G) = 0$ for all m . Alternatively, the contract could have subjective performance pay $c(m, r_B) = 0$ and $c(m, r_G) > 0$, coupled with $b(m = \underline{m}) = 0$ and $b(m \neq \underline{m}) = -c(m, r)$ for all m and r .²³ Given that the agent receives the same

²³Note that for both types of contract, the agent's limited liability constraint implies that

bonus for meeting both the performance target and the supervisor's request, both contracts provide the same incentives at equal cost. In the analysis below, we use the latter type of contract, which is without loss of generality. For ease of notation, below we denote $c(m, r_G) = c$.

Before deriving the optimal contract, we first establish that not all possible effort allocations can be implemented through the combination of objective and subjective performance pay.

Lemma 2 *If $\eta > \varphi$ ($\eta < \varphi$), effort allocations such that $\frac{e_2}{e_1} > \frac{1-\varphi}{\varphi}$ ($\frac{e_2}{e_1} < \frac{1-\varphi}{\varphi}$) cannot be implemented.*

Proof. In this proof, we focus on the case where $\eta > \varphi$. The case $\eta < \varphi$ is the mirror image. Consider any combination of (subjective) performance pay c and performance target \underline{m} by the principal, and requested efforts \underline{e}_1 and \underline{e}_2 by the supervisor, such that the agent receives the bonus unless $r = r_B$ or $m \neq \underline{m}$. Several outcomes are possible. First, the agent can ignore both the principal's target and the supervisor's request, hoping to be able to manipulate m into \underline{m} . Then, optimal effort is zero. Second, if the agent ignores the supervisor's request but adheres to the principal's target $m = \underline{m}$, the agent's optimal effort allocation has $\frac{e_2}{e_1} = \frac{1-\varphi}{\varphi}$ (see (PMR)). Third, the supervisor can ignore the principal's target and induce the agent to follow her request. When ignoring the target, the supervisor's optimal effort allocation is such that $\frac{e_2}{e_1} = \frac{1-\eta}{\eta} < \frac{1-\varphi}{\varphi}$ (see (SR)), where the inequality sign follows from $\eta > \varphi$. Lastly, the agent can follow the supervisor's request while the supervisor adheres to the principal's target. In the latter situation, the supervisor's demands must meet the agent's incentive compatibility constraint. Using the agent's utility function (2.2), this constraint is given by:

$$c - \frac{1}{2}\underline{e}_1^2 - \frac{1}{2}\underline{e}_2^2 \geq \max \left\{ (1-p)(1-q)c, (1-p)c - \frac{1}{2} \frac{1}{\varphi^2 + (1-\varphi)^2} \underline{m}^2 \right\} \quad (2.21)$$

total compensation can never be negative.

The first term in braces is the agent's expected utility when exerting no effort at all. The second term in braces is the agent's expected utility when reaching $m = \underline{m}$ at lowest effort cost as given by effort levels (2.5) and (2.6), i.e. by choosing the performance measure effort ratio (PMR). It follows that if the agent can be induced to exert any effort allocation on $m = \underline{m}$, an effort allocation such that $\frac{e_2}{e_1} = \frac{1-\varphi}{\varphi}$ is also feasible. A straightforward extension of the analysis in Section 2.3.2 yields that the principal can implement effort allocations satisfying $\frac{e_2}{e_1} = \frac{1-\varphi}{\varphi}$ by setting $c = \frac{1}{2} \frac{(p+(1-p)q)\underline{m}^2}{\varphi^2+(1-\varphi)^2}$. Keeping \underline{m} constant, an increase in c makes different effort allocations on $m = \underline{m}$ feasible. As $\eta > \varphi$, the supervisor prefers effort allocations on $m = \underline{m}$ with more e_1 over effort allocations with less e_1 . Hence, the supervisor optimally requests an effort allocation on \underline{m} such that $\frac{e_2}{e_1} \leq \frac{1-\varphi}{\varphi}$. Summarizing, none of these outcomes has $\frac{e_2}{e_1} > \frac{1-\varphi}{\varphi}$. ■

Lemma 2 is illustrated by Figure 2.2, for the case where $\eta > \varphi$. The figure depicts the $m = \underline{m}$ performance target which runs orthogonal to the line representing the performance measure's effort ratio (PMR), as well as the set of points where the agent's incentive compatibility constraint (ICC) given by (2.21) is binding. The supervisor can only implement effort allocations on the $m = \underline{m}$ line below the ICC, as indicated by the thicker line segment. This line segment always includes the effort ratio on the line (PMR), as this is the effort ratio that yields $m = \underline{m}$ at minimal effort cost. The supervisor's indifference lines run orthogonal to the line representing the supervisor's preferred effort ratio (SR). If $\eta > \varphi$, these indifference lines are steeper than the $m = \underline{m}$ line. Hence, the supervisor prefers to implement an effort allocation given by the intersection between the $m = \underline{m}$ line and the ICC with the highest e_1 . This implies that the principal cannot induce the supervisor to request an effort allocation that lies beyond line (PMR) from the supervisor's perspective, as given $m = \underline{m}$ and $\eta > \varphi$ the supervisor always prefers an effort ratio $\frac{e_2}{e_1} = \frac{1-\varphi}{\varphi}$ over effort ratio's such that $\frac{e_2}{e_1} > \frac{1-\varphi}{\varphi}$. Similarly, if $\eta < \varphi$, the supervisor always prefers to induce effort ratio $\frac{e_2}{e_1} = \frac{1-\varphi}{\varphi}$ over effort ratio's such that $\frac{e_2}{e_1} < \frac{1-\varphi}{\varphi}$.²⁴

²⁴Note that it is always optimal for the supervisor to request effort levels \underline{e}_i that the agent

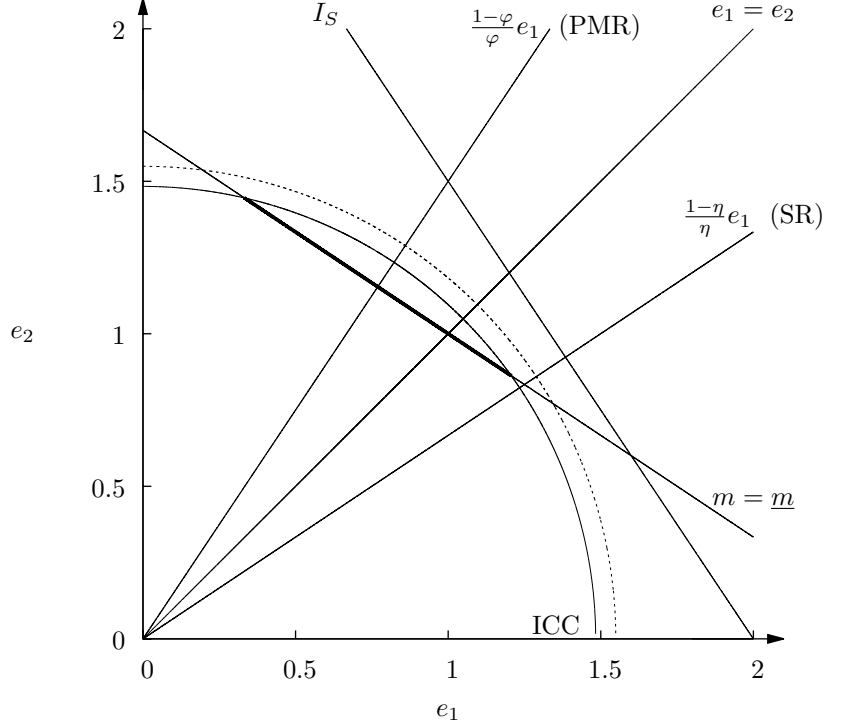


Figure 2.2: An example with $\eta = \frac{3}{5}$ and $\varphi = \frac{2}{5}$. The feasible set of demands for the supervisor is given by the bold chord, which is always centered around (PMR). Lowering \underline{m} or raising the ICC can lengthen this chord.

Lemma 2 underlines an important difference between objective and subjective performance evaluation. If the supervisor would be a second verifiable performance measure with bias $\eta \neq \varphi$, any effort allocation would have been feasible, as shown by Feltham and Xie (1994). This holds even when the two measures are biased in the same direction. By punishing too high performance on the most-biased measure combined with rewarding high performance on the least-biased measure, the agent could be induced to choose effort allocations that are less biased than the least-biased performance measure. Here, however, the supervisor can use her discretionary power to make the agent choose an effort allocation closer to her optimal ratio, as she subjectively determines whether the agent's

accepts (see (2.21)), such that in equilibrium, we always obtain $e_i = \underline{e}_i$.

performance is good or bad. Hence, if the supervisor is more biased than the verifiable performance measure ($\eta > \varphi > \frac{1}{2}$ or $\frac{1}{2} > \varphi > \eta$), the supervisor uses any leeway to increase the distortion in efforts. It follows that in this situation, the least-biased effort ratio that can be implemented is given by the performance measure's effort ratio (PMR), $\frac{e_2}{e_1} = \frac{1-\varphi}{\varphi}$.

It is possible to implement effort allocations beyond the supervisor's preferred effort ratio (SR), i.e. to implement $\frac{e_2}{e_1} < \frac{1-\eta}{\eta}$ when $\eta > \varphi$ and vice versa, although it requires commitment to penalize too high verifiable performance. This can be easily inferred from Figure 2.2: by raising c , the ICC can be shifted up such that the intersection between the $m = \underline{m}$ line and the agent's ICC lies to the right of the line representing the supervisor's preferred effort ratio (SR). Then, the supervisor would prefer to request the effort allocation given by the intersection of the ICC with (SR). However, this would generate measured performance $m > \underline{m}$, implying that with probability q the principal would not pay out the bonus even when the agent would follow the supervisor's request. Anticipating this, the agent would not accept the supervisor's request, given that q is sufficiently high. The best feasible effort allocation for the supervisor is again where the ICC intersects the $m = \underline{m}$ line with highest e_1 , which implies that $\frac{e_2}{e_1} < \frac{1-\eta}{\eta}$.

Figure 2.2 also helps to build intuition for the results that follow. A reliable performance measure (high q) allows the principal to enforce a particular level of objectively measured performance m at low cost. This restricts the supervisor, who cannot simply implement an effort allocation with her most-preferred effort ratio (SR), as was possible in the absence of the verifiable performance measure (Section 2.3.3). Yet, the performance measure also helps imperfect supervisors by increasing the chance that a shirking agent is caught. This allows supervisors to request higher effort levels for a given discretionary bonus c .

If monitoring or performance measurement is sufficiently imperfect, two additional constraints arise. First, if the supervisor's effectiveness is sufficiently low, it becomes attractive for the agent to only pretend to abide by the supervisor's request. Instead, the agent could obtain the same level of measured performance

$m = m(\underline{e}_1, \underline{e}_2)$ at lower cost by choosing efforts along the performance measure's effort ratio (PMR). Second, if the performance measure is sufficiently unreliable (low q), the supervisor is tempted to ignore the principal's target $m = \underline{m}$ and request effort levels along her preferred effort ratio (SR). The optimal contract depends on whether these constraints are binding.

Propositions 1 to 4 below give the optimal contract, differentiated by whether or not these constraints are binding. We focus on the case where $\eta > \varphi$, the case of $\eta < \varphi$ is the mirror image.

Proposition 1 *Given $\eta > \varphi$, the optimal contract induces balanced effort $e_1 = e_2$ if and only if (i) $\varphi \leq \frac{1}{2}$ and (ii) $\frac{1}{(1-2\eta)^2} \frac{q}{(1-q)} \geq p \geq \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$.*

Under these conditions, the optimal contract has $c = \underline{m} = (p + (1-p)q)$, $b(m = \underline{m}) = 0$, and $b(m \neq \underline{m}) = -c$.

The equilibrium effort levels are $e_1 = e_2 = (p + (1-p)q)$.

Proof. The proof is given in the Appendix. ■

Proposition 1 gives the necessary and sufficient conditions under which the principal can induce balanced efforts. The outcome is as if the principal has access to an unbiased performance measure that is as effective as the supervisor and the verifiable measure combined. Condition (i) follows from Lemma 2: the principal cannot induce the supervisor to implement balanced efforts when the supervisor is more biased than the verifiable performance measure. Condition (ii) implies that the effectiveness of the supervisor and the performance measure should not be too different. If the supervisor's effectiveness p is too low, the agent is tempted to feign satisfying the supervisor's demands by generating $m = m(\underline{e}_1, \underline{e}_2)$ at lower cost along the PMR. The supervisor cannot use the objective performance measure to detect this breach of the implicit agreement. This is more attractive to the agent when the effectiveness of the performance measure (supervisor) is larger (smaller). The benefits of this deviation for the agent are larger when the performance measure is more biased. In contrast, if the performance measure is relatively unreliable, the supervisor would ignore the performance target \underline{m} . If

the probability that the agent can manipulate measured performance is high, the principal is unlikely to detect the deviation. This increases the incentive of the agent to follow the supervisor in deviating. Deviating is more beneficial for more biased supervisors.

The following Propositions describe the optimal outcomes when the conditions in Proposition 1 are not met. Proposition 2 considers the case where $\varphi > \frac{1}{2}$, while Propositions 3 and 4 consider the cases where p and q are too different.

Proposition 2 *Given $\eta > \varphi > \frac{1}{2}$, the principal optimally induces efforts along the PMR ($\frac{e_2}{e_1} = \frac{1-\varphi}{\varphi}$), if and only if $p \leq \frac{q}{(1-q)} \frac{(\eta\varphi+(1-\eta)(1-\varphi))^2}{(\eta-\varphi)^2}$.*

Under these conditions, the optimal contract has $c = \frac{(p+(1-p)q)}{2(\varphi^2+(1-\varphi)^2)}$,

$\underline{m} = (p + (1-p)q)$, $b(m = \underline{m}) = 0$, and $b(m \neq \underline{m}) = -c$.

The equilibrium effort levels are $e_1 = (p + (1-p)q) \frac{\varphi}{(\varphi^2+(1-\varphi)^2)}$ and $e_2 = (p + (1-p)q) \frac{1-\varphi}{(\varphi^2+(1-\varphi)^2)}$.

Proof. The proof is given in the Appendix. ■

Figure 2.3 gives an example of this situation. Given $\eta > \varphi > \frac{1}{2}$, the supervisor uses any available room to move away from the PMR to distort the agent's efforts even further away from the principal's objectives. Hence, given that it is not possible for the principal to induce more balanced effort levels, it is optimal to constrain the supervisor to the PMR. Compared to the case where $\varphi \leq \frac{1}{2}$, the principal demands the same measured performance \underline{m} , but sets a smaller bonus. Thereby, the principal makes it impossible for the supervisor to request effort levels that are more biased than the verifiable performance measure. Even though this implies that the supervisor does not mitigate the bias inherent in the objective performance measure, she still adds value for the principal. The supervisor provides additional monitoring, which allows for a reduction in the rents left to the agent. The resulting outcome is as if the principal has access to a biased performance measure as effective as the supervisor and the verifiable measure combined. The only constraint on the outcome is that the supervisor should not be tempted to ignore the target. As before, this happens when the

verifiable measure is sufficiently unreliable relative to the effectiveness of the supervisor, and is more likely when the PMR and the SR are more apart.²⁵

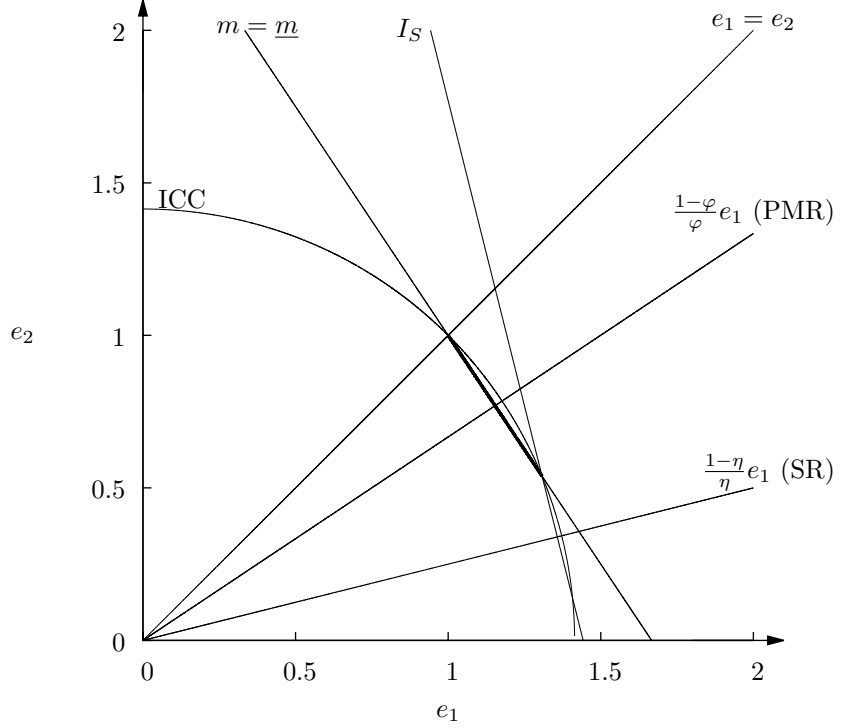


Figure 2.3: An example with $\eta = \frac{4}{5}$ and $\varphi = \frac{3}{5}$. The feasible set of demands for the supervisor is given by the bold chord, which is always centered around (PMR). Even though balanced efforts are feasible here the supervisor prefers to induce highly distorted efforts. Hence, the principal is better off raising \underline{m} to the intersection between (PMR) and the ICC.

Combining Propositions 1 and 2, it follows that if p and q are sufficiently large, the supervisor's bias only affects whether balanced efforts can be implemented. Other than that, neither the optimal contract nor equilibrium efforts depend on the bias of the supervisor. The following proposition describes the optimal contract when the supervisor is (relatively) ineffective.

²⁵As the contract in Proposition 2 implements effort levels that meet the principal's performance target \underline{m} at minimal effort costs, the agent never benefits from choosing different effort levels while still generating measured performance \underline{m} .

Proposition 3 Suppose $\eta > \varphi$ and $\varphi \leq \frac{1}{2}$. If $p < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, the optimal contract induces unbalanced efforts biased towards the PMR, $\frac{e_2}{e_1} > 1$.

Under these conditions, the optimal contract has $c = \frac{(\sqrt{(1-p)q} + (1-2\varphi)\sqrt{p})^2}{2(\varphi^2 + (1-\varphi)^2)}$, $\underline{m} = (1-p)q + (1-2\varphi)\sqrt{p(1-p)q}$, $b(m = \underline{m}) = 0$, and $b(m \neq \underline{m}) = -c$.

The equilibrium effort levels are $e_1 = \frac{(\sqrt{(1-p)q} + (1-2\varphi)\sqrt{p})}{\varphi^2 + (1-\varphi)^2} (\varphi\sqrt{(1-p)q} + (1-\varphi)\sqrt{p})$ and $e_2 = \frac{(\sqrt{(1-p)q} + (1-2\varphi)\sqrt{p})}{\varphi^2 + (1-\varphi)^2} ((1-\varphi)\sqrt{(1-p)q} - \varphi\sqrt{p})$.

Proof. The proof is given in the Appendix. ■

If the supervisor is ineffective, the agent is tempted to feign compliance by meeting the performance target $m = \underline{m}$ at minimal effort cost rather than by following the supervisor's request. The optimal deviation would be to choose the effort ratio equal to the performance measure's ratio as given by (PMR). Feigning compliance is particularly attractive for the agent when the performance measure is both highly effective and highly biased while the supervisor is weak. Anticipating the agent's incentive to deviate, the supervisor is forced to shift her requested effort levels closer to the PMR. This increases the agent's rents, as can be seen from Figure 2.2.

The principal optimally responds to this inefficiency in two ways. First, he sets a higher performance target \underline{m} as compared to the case with a more effective supervisor. While this forces the supervisor to request a biased effort ratio closer to the PMR, it also reduces the agent's rents when feigning compliance, which makes following the supervisor's request relatively more attractive. Second, as the implemented effort levels remain biased toward the PMR, the principal lowers the bonus. Hence, compared to the case with a more effective supervisor (Proposition 1), the bonus is smaller while the performance target is higher.

The imbalance in the effort levels is decreasing in supervisor effectiveness p and increasing in the bias of the verifiable performance measure $|\frac{1}{2} - \varphi|$.²⁶ Provided that the supervisor is not more biased than the verifiable performance measure ($\eta > \varphi$ and $\varphi \leq \frac{1}{2}$), the supervisor's bias is irrelevant. For the principal, this

²⁶Note that if $p = 0$, the equilibrium is identical to the equilibrium derived in subsection 3.2 where only a verifiable performance measure was available.

implies that if only weak supervisors are available, their bias is not a concern. This differs when supervisors are relatively strong, as shown in Proposition 4.

Proposition 4 *Suppose $\eta > \varphi$. If $p > \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$ and $\varphi \leq \frac{1}{2}$, or if $p > \frac{q}{(1-q)} \frac{(\eta\varphi + (1-\eta)(1-\varphi))^2}{(\eta-\varphi)^2}$ and $\varphi > \frac{1}{2}$, the optimal contract induces unbalanced efforts biased towards the SR.*

Under these conditions, the optimal contract has $c = \frac{(\sqrt{p(1-q)} + |1-2\eta|\sqrt{q})^2}{2(\eta^2 + (1-\eta)^2)}$,

$b(m = \underline{m}) = 0$, and $b(m \neq \underline{m}) = -c$.

If $\eta > \frac{1}{2}$, $\underline{m} = \lambda \left((\eta\varphi + (1-\eta)(1-\varphi)) \sqrt{p(1-q)} + (\eta-\varphi) \sqrt{q} \right)$ and equilibrium effort levels are $e_1 = \lambda \left(\eta \sqrt{p(1-q)} - (1-\eta) \sqrt{q} \right)$ and

$e_2 = \lambda \left((1-\eta) \sqrt{p(1-q)} + \eta \sqrt{q} \right)$, where $\lambda = \frac{(\sqrt{p(1-q)} + |1-2\eta|\sqrt{q})}{(\eta^2 + (1-\eta)^2)}$.

If $\eta < \frac{1}{2}$, $\underline{m} = \lambda \left((\eta\varphi + (1-\eta)(1-\varphi)) \sqrt{p(1-q)} - (\eta-\varphi) \sqrt{q} \right)$ and equilibrium effort levels are $e_1 = \lambda \left(\eta \sqrt{p(1-q)} + (1-\eta) \sqrt{q} \right)$ and

$e_2 = \lambda \left((1-\eta) \sqrt{p(1-q)} - \eta \sqrt{q} \right)$.

Proof. The proof is given in the Appendix. ■

When the verifiable performance measure is unreliable, strong supervisors have an incentive to ignore the performance target \underline{m} and deviate to their most-preferred effort ratio (SR). This reduces the agent's incentives, as it implies that the agent's bonus is forfeit if they cannot manipulate measured performance. However, if this probability q is small enough, the supervisor is still better off inducing the agent to allocate efforts along the SR. Figure 2.4 depicts this situation. The solid ICC curve gives the effort allocations the supervisor could induce provided that the penalty $b(m \neq \underline{m})$ would never be incurred, while the dashed ICC gives all implementable effort allocations if penalty $b(m \neq \underline{m})$ is incurred with probability q . From the supervisor's perspective, the best feasible effort allocation implementing $m = \underline{m}$ is the rightmost point on the thick segment of the $m = \underline{m}$ line. However, given that the supervisor is very effective compared to the verifiable measure, she can induce efforts off the $m = \underline{m}$ line that yield her higher utility, as depicted by the shaded area. Optimally, she would induce the agent to choose the effort levels determined by the intersection of the SR and the

dashed ICC.

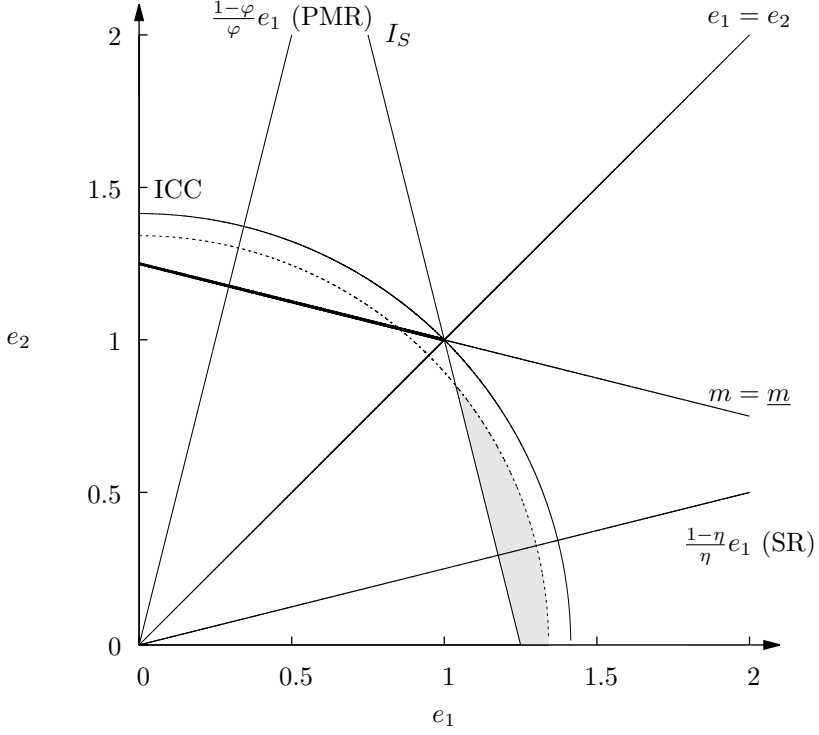


Figure 2.4: An example with $\eta = \frac{4}{5}$ and $\varphi = \frac{1}{5}$. For relatively low q , it is feasible and optimal for the supervisor to ignore $m = \underline{m}$, as shown by the shaded area.

Anticipating the supervisor's incentive to deviate, the principal optimally adjusts the contract. First, given bonus c , the principal adjusts the performance target \underline{m} such that the supervisor can induce the agent to meet $m = \underline{m}$ with an effort allocation closer to SR. Given that $\eta > \varphi$, this adjustment is upward (downward) when $\eta < \frac{1}{2}$ ($\eta > \frac{1}{2}$). Second, as the implemented effort ratio is distorted, the principal optimally sets a smaller bonus.

Proposition 4 shows that an unreliable performance measure not only hampers the provision of incentives to the agent, but also makes monitoring the supervisor difficult. Strong supervisors use their information advantage to induce efforts that are biased towards their most-preferred task. To mitigate this problem, the

principal must accommodate to the supervisor's preferences.²⁷ Given q , the imbalance in equilibrium efforts is increasing in p . However, this does not imply that the principal would prefer a less effective supervisor when verifiable performance measures are unreliable. Effective supervisors also reduce the agent's rents, which in turn makes implementing higher efforts more attractive to the principal. This positive effect on efficiency dominates the negative effect of a more biased outcome. Given supervisor effectiveness p , an increase in the supervisor's bias harms the principal.

The following proposition gives an overview of the comparative statics of all parameters on the optimal level of bonus pay and on the principal's payoff, for all cases considered above combined. As our linear-quadratic framework implies that the principal's payoff is equal to the agent's bonus pay, these comparative statics are identical.

Proposition 5 *Given $\eta > \varphi$, comparative statics on the optimal level of bonus pay and the principal's equilibrium payoff are as follows:*

- (i) $\frac{\partial U_p}{\partial p} = \frac{\partial c}{\partial p} \geq 0$, with equality only if $q = 1$.
- (ii) $\frac{\partial U_p}{\partial q} = \frac{\partial c}{\partial q} \geq 0$, with equality only if $p = 1$.
- (iii) $\frac{\partial U_p}{\partial \varphi} = \frac{\partial c}{\partial \varphi} \geq 0$ if $\varphi < \frac{1}{2}$, with equality only if $p \geq \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, and $\frac{\partial U_p}{\partial \varphi} = \frac{\partial c}{\partial \varphi} \leq 0$ if $\varphi > \frac{1}{2}$, with equality only if $p > \frac{(\eta\varphi+(1-\eta)(1-\varphi))^2}{(\eta-\varphi)^2} \frac{q}{(1-q)}$.
- (iv) $\frac{\partial U_p}{\partial \eta} = \frac{\partial c}{\partial \eta} \geq 0$ if $\eta < \frac{1}{2}$, with equality only if $p \leq \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$, and $\frac{\partial U_p}{\partial \eta} = \frac{\partial c}{\partial \eta} \leq 0$ if $\eta > \frac{1}{2}$, with equality only if $\varphi < \frac{1}{2}$ and $p \leq \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$ or if $\varphi > \frac{1}{2}$ and $p < \frac{(\eta\varphi+(1-\eta)(1-\varphi))^2}{(\eta-\varphi)^2} \frac{q}{(1-q)}$.

Proof. The proof is given in the Appendix. ■

To summarize our findings above, we find that the principal always benefits from more effective performance measurement and supervision. Both reduce the agent's rents and allow for stronger incentive pay. Supervisor ability and reliability of the verifiable measure are substitutes. A biased supervisor need

²⁷Note that if $q = 0$, the equilibrium is identical to the equilibrium derived in subsection 3.3 where only subjective performance evaluation was available. For $q = 0$, the (irrelevant) performance target \underline{m} simply gives the performance as measured given the agent's efforts.

not be detrimental to the principal, provided that the principal has access to a verifiable performance measure that is sufficiently effective and the supervisor is either less biased than the verifiable performance measure or biased towards the other task. However, when the verifiable performance measure is too unreliable, a biased supervisor forces the principal to accommodate to her preferences, leading to lower optimal incentive pay. Similarly, a biased performance measure is not problematic as long as the supervisor is sufficiently effective. If not, the agent's incentive to ignore the supervisor leads to an allocation of efforts biased towards the task that receives most weight in the performance measure.

We have not considered bonus pay for the principal. Supervisor's bonus pay could only be based on the verifiable performance measure m . Effectively, this would draw the supervisor's relative preferences over tasks closer to PMR. From Proposition 5, it follows that this benefits the principal only when the performance measure is relatively weak and less biased or biased towards the opposite task as the supervisor. However, given the supervisor's limited liability, providing a bonus to the supervisor is highly costly precisely when the verifiable measure is weak. Hence, incentive pay for the supervisor based on the agent's performance is of little use to the principal.

The following proposition establishes that the principal is always better off combining subjective and objective performance evaluation compared to using only subjective or only objective performance evaluation. Combined evaluation always reduces the agent's rents, which benefits the principal. Furthermore, if both subjective and objective performance measurement are sufficiently effective, as studied in Propositions 1 and 2, combined evaluation can eliminate the bias that might arise when only using objective or subjective performance evaluation. On the other hand, if p and q are sufficiently different, combined evaluation might increase the bias as compared to using only objective or only subjective performance evaluation. Still, for all performance measures and all supervisors, the cost implied by the extra bias is outweighed by the benefits that arise from

better monitoring of the agent.²⁸

Proposition 6 *The principal always benefits from combining subjective and objective performance evaluation.*

Proof. The proof is given in the Appendix. ■

2.5 Generalists versus specialists as supervisors

We use our results to discuss the principal's trade-off when choosing between different supervisors. In particular, we consider the choice between a specialist and a generalist. We assume that a specialist is more effective in monitoring the agent than a generalist, $p_S > p_G$, where subscript S (G) denotes specialist (generalist). However, the specialist has stronger preferences regarding the execution of the tasks than the generalist. We assume that the generalist has unbiased preferences, $\eta_G = \frac{1}{2}$, while the specialist has biased preferences $\eta_S \neq \frac{1}{2}$. This combination of expertise and bias corresponds to findings by Li (2013), who looks at decision made by reviewers of grant proposals at the US National Institute of Health. She finds that reviewers are better able to judge the quality of proposals in their own area, but that they are also biased in favour of proposals in their own area. While the relation between grant reviewers and potential grant recipients differs from the relation between managers and employees, the reviewers decide about resource allocation across different fields, much like the supervisor in our model.

To determine whether the principal prefers a generalist or a specialist in the absence of verifiable performance measures, we can directly use the principal's payoff under pure discretionary pay (2.19). It follows that the principal prefers the specialist manager if

$$\frac{p_S}{p_G} > 2 \left((1 - \eta_S)^2 + \eta_S^2 \right) \quad (2.22)$$

²⁸We have assumed that the (fixed) cost of obtaining both the objective and the subjective performance measure is zero. If these costs are positive, a trade-off naturally arises.

Hence, in choosing between a specialist and a generalist, the principal faces a trade-off between more effective supervision and more distorted effort levels. The principal tolerates the bias of the specialist only if she is sufficiently more effective than the generalist.

Now suppose that the principal has access to a verifiable performance measure. Without loss of generality, we focus on the case where $\eta_S > \frac{1}{2}$ and $\eta_S > \varphi$. Better performance measurement (higher q) affects the choice between the generalist and the specialist in two ways. First, the marginal benefit to the principal of a better supervisor is smaller, as the effectiveness of subjective and objective performance measurement are substitutes. Second, the principal can use the verifiable performance measure to neutralize or at least mitigate the bias that is induced by the specialist supervisor. The latter effect dominates, unless the specialist's bias cannot be fully eliminated.

Combining Propositions 1, 3, and 5, we have that if $\varphi \leq \frac{1}{2}$ and $p_S < \frac{1}{(1-2\eta_S)^2} \frac{q}{(1-q)}$ the supervisor's bias is irrelevant while the principal benefits from more effective supervision. It follows that the principal prefers the specialist: when the verifiable performance measure is sufficiently reliable and biased towards the opposite task as compared to the specialist, the principal can neutralize the specialist's bias while obtaining the benefits from better supervision. If the objective measure is relatively unreliable, the principal is forced to accommodate towards the specialist's bias. If the bias of the specialist is sufficiently strong while she is only slightly more effective, the principal prefers the generalist (in the limit where $q = 0$, this condition is given by (2.22)). Interestingly, if $\varphi > \frac{1}{2}$ and the verifiable measure is sufficiently reliable, the bias induced by the specialist is equal to the bias in the verifiable performance measure. In this case, a more aligned objective performance measure (φ closer to $\frac{1}{2}$) makes the specialist more attractive, whereas a more reliable objective performance measure (higher q) increases the relative attractiveness of the generalist. These results are summarized in the following proposition.

Proposition 7 *Consider the principal's choice of supervisor between a specialist and a generalist, where $\eta_S > \eta_G = \frac{1}{2}$ and $p_S > p_G$. The principal prefers the specialist over the generalist, unless (i) $\eta_S > \varphi > \frac{1}{2}$ and $p_S < \frac{(\eta_S \varphi + (1-\eta_S)(1-\varphi))^2}{(\eta_S - \varphi)^2} \frac{q}{(1-q)}$ or (ii) $p_G > \frac{1}{(1-2\eta_S)^2} \frac{q}{(1-q)}$. In case (i), if $p_G \geq \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, the principal prefers the generalist if and only if $p_S - p_G < (1-2\varphi)^2 \left(\frac{q}{(1-q)} + p_G \right)$, while if $p_G < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, the principal prefers the generalist if and only if $p_S - p_G < \frac{2\sqrt{p_G}}{(1-q)} \left((2\varphi - 1) \sqrt{(1-p_G)q} - 2\varphi(1-\varphi) \sqrt{p_G} \right)$. In case (ii), the principal prefers the generalist if and only if $p_S - p_G < \frac{((2\eta-1)\sqrt{p_S} - \sqrt{\frac{q}{(1-q)}})^2}{2(\eta^2 + (1-\eta)^2)}$.*

Proof. The proof is given in the Appendix. ■

2.6 Concluding remarks

We have studied the effects of biased supervision in a three-tier hierarchy. Supervisors can influence their subordinates' effort allocation across tasks by (ab)using their discretion in determining subjective performance evaluations. This allows supervisors to direct their subordinates towards activities that have relatively high value for the supervisor but not necessarily for the organization. Biased supervision is detrimental for organizations in the absence of other performance measures, as it leads to misallocation of agents' effort across tasks. However, this negative effect of supervisor bias is mitigated when the principal can also use a verifiable performance measure, even when the latter is highly incongruent. A biased supervisor will use her discretionary powers to direct her agents away from the distortion inherent in the verifiable performance measure, towards tasks she considers more important. At the same time, the principal can use verifiable performance targets to constrain the supervisor. We have shown that the optimal level of incentive pay for employees is decreasing in supervisor bias only when the verifiable performance measure is unreliable.

We have derived the optimal contract assuming that the supervisor's bias and ability are observed by the principal. If the supervisor's type is unobservable

Appendix

and supervisors self-select into organizations, a given contract is most attractive to supervisors with high-ability and with a bias close to the bias of the verifiable performance measure. This implies that in determining performance targets and the supervisor's discretion, the principal faces a trade-off between attracting aligned but low-ability supervisors and attracting high-ability but more biased supervisors. The effects of (incentive) wages on the self-selection of workers based on intrinsic motivation and/or ability is studied by e.g. Handy and Katz (1998), Besley and Ghatak (2005), Delfgaauw and Dur (2008, 2010), Prendergast (2007), and Dal Bo et al. (2013). How self-selection of managers is affected by the degree of discretion over their subordinates' activities is an interesting question that we leave for future work.

2.A Appendix

Proof of Lemma 1. Suppose the principal wants to induce some efforts $e^* = \{e_1^*, e_2^*\}$, which would lead to measured performance $m = \underline{m}$. First, suppose the supervisor also wants to induce e^* . Suppose that agent exerts effort e^* . Provided that $w_A(\underline{m}, r_G) \geq w_A(m \neq \underline{m}, r)$, the agent will not manipulate m ex post. Hence, expected utility is given by

$$U_A(e^*) = w_A(\underline{m}, r_G) - \frac{1}{2} (e_1^*)^2 - \frac{1}{2} (e_2^*)^2 \quad (\text{A1})$$

The agent's rents (the difference between wage $w_A(\underline{m}, r_G)$ and the agent's effort cost) are determined by $\max_e U_A(e^-)$, where $U_A(e^-)$ represents the agent's expected utility after choosing any efforts $e \neq e^*$. Consider any $e \neq e^*$ such that $m \neq \underline{m}$. Now, the agent will manipulate m when given the option, implying that

$$U_A(e^-) = qw_A(m \neq \underline{m}, r_B) + p(1 - q)w_A(\underline{m}, r_B) + (1 - p)(1 - q)w_A(\underline{m}, r_G) - \frac{1}{2} (e_1^-)^2 - \frac{1}{2} (e_2^-)^2 \quad (\text{A2})$$

Note that with probability $(1-p)q$ the supervisor send report r_B even though she did not observe the agent's efforts, because she learned from (non-manipulated) measured performance that the agent did not adhere to her demands. Lastly, for any $e \neq e^*$ such that $m = \underline{m}$, we have

$$U_A(e^{-*}) = pw_A(\underline{m}, r_B) + (1-p)w_A(\underline{m}, r_G) - \frac{1}{2}(e_1^{-*})^2 - \frac{1}{2}(e_2^{-*})^2$$

It follows that the agent's rents $\max_e U_A(e^{-*})$ are minimal when $w_A(m \neq \underline{m}, r_B) = w_A(\underline{m}, r_B) = 0$.

Next, consider the supervisor's incentive to demand $e^S \neq e^*$. First, consider any e^S yielding $m = \underline{m}$. In terms of measured performance, the principal cannot distinguish between e^S and e^* . As a result, the effect of the wage scheme on the agent's incentive to exert e^S is identical to the incentive to exert e^* as discussed above. Hence, if effort cost at e^* are at least as large as effort cost at e^S , the principal cannot prevent the supervisor from inducing e^S .

Second, consider any e^S yielding $m = m^S \neq \underline{m}$. Now, the agent might optimally manipulate measured performance into either m^S or \underline{m} . Suppose first that the (supervisor demands that the) agent manipulates m into \underline{m} when the opportunity arises. If the agent follows the supervisor's request, $e = e^S$, his expected utility equals

$$U_A(e^S) = qw_A(m \neq \underline{m}, r_G) + (1-q)w_A(\underline{m}, r_G) - \frac{1}{2}(e_1^S)^2 - \frac{1}{2}(e_2^S)^2 \quad (\text{A3})$$

Now suppose the agent exerts $e \neq e^S$ such that $m \neq m^S$ and $m \neq \underline{m}$. This yields expected utility

$$U_A(e \neq e^S) = qw_A(m \neq \underline{m}, r_B) + p(1-q)w_A(\underline{m}, r_B) + (1-p)(1-q)w_A(\underline{m}, r_G) - \frac{1}{2}(e_1)^2 - \frac{1}{2}(e_2)^2 \quad (\text{A4})$$

Appendix

Next, consider $e \neq e^S$ such that $m = \underline{m} \neq m^S$. This yields

$$U_A(e \neq e^S) = pw_A(\underline{m}, r_B) + (1-p)w_A(\underline{m}, r_G) - \frac{1}{2}(e_1)^2 - \frac{1}{2}(e_2)^2$$

Lastly, the agent can exert $e \neq e^S$ such that $m = m^S$, which yields

$$\begin{aligned} U_A(e \neq e^S) &= pqw_A(m \neq \underline{m}, r_B) + p(1-q)w_A(\underline{m}, r_B) + (1-p)qw_A(m \neq \underline{m}, r_G) + \\ &\quad + (1-p)(1-q)w_A(\underline{m}, r_G) - \frac{1}{2}(e_1)^2 - \frac{1}{2}(e_2)^2 \end{aligned}$$

The set of e^S the supervisor can demand increases in the difference between $U_A(e^S)$ and $\max_e U_A(e \neq e^S)$. Hence, the principal wants to minimize this difference. It is not possible to set $w_A(m \neq \underline{m}, r_B) > w_A(m \neq \underline{m}, r_G)$, as the cheap talk nature of the reports implies that the supervisor would increase the set e^S by switching the report labels. Hence, it is optimal for the principal to set $w_A(m \neq \underline{m}, r_G) = w_A(m \neq \underline{m}, r_B) = 0$. It follows that the agent's wage is optimally zero if $m \neq \underline{m}$, regardless of the supervisor's report. Furthermore, $w_A(m \neq \underline{m}, r_G) = w_A(m \neq \underline{m}, r_B) = 0$ also implies that manipulation of m into m^S is never beneficial (for both the agent and the supervisor).

Lastly, we show that $w_A(\underline{m}, r_B) > 0$ does not affect the supervisor's decision to induce e^S rather than e^* , implying that optimally $w_A(\underline{m}, r_B) = 0$ to minimize the agent's rents as derived above. We focus on the case where the agent's best alternative to following the supervisor's demand is an effort allocation such that $m \neq \underline{m}$. The case where the best alternative yields $m = \underline{m}$ is analog and therefore omitted. If the agent decides not to follow the supervisor's demands (either e^* or e^S), it follows from substituting for $w_A(m \neq \underline{m}, r_G) = w_A(m \neq \underline{m}, r_B) = 0$ into (A2) and (A4) that exerting $e_1 = e_2 = 0$ is optimal. Using (A1), it follows that to demand e^* , $w_A(\underline{m}, r_G)$ and $w_A(\underline{m}, r_B)$ have to be such that

$$(p(1-q) + q)w_A(\underline{m}, r_G) - \frac{1}{2}(e_1^*)^2 - \frac{1}{2}(e_2^*)^2 \geq p(1-q)w_A(\underline{m}, r_B) \quad (\text{A5})$$

Both the principal and the supervisor are best off when these equations hold with equality. Suppose that e^* is such that $\frac{e_2^*}{e_1^*} = \frac{1-\kappa}{\kappa}$, for any $\kappa \in [0, 1]$. By (A5), this implies that $e_1^* = \kappa \sqrt{2(p(1-q) + q) w_A(\underline{m}, r_G) - 2p(1-q)w_A(\underline{m}, r_B)}$ and $e_2^* = \frac{1-\kappa}{\kappa} e_1^*$. Using supervisor's utility (2.4), a supervisor with bias η derives utility from inducing e^* equal to

$$U_S(e^*) = (\eta\kappa + (1-\eta)(1-\kappa)) \sqrt{2(p(1-q) + q) w_A(\underline{m}, r_G) - 2p(1-q)w_A(\underline{m}, r_B)}$$

Similarly, using (A3), it follows that the supervisor can demand e^S if

$$p(1-q)w_A(\underline{m}, r_G) - \frac{1}{2} (e_1^S)^2 - \frac{1}{2} (e_2^S)^2 \geq p(1-q)w_A(\underline{m}, r_B)$$

By (SR), a supervisor with bias η will optimally induce the agent to exert efforts such that $\frac{e_2^S}{e_1^S} = \frac{1-\eta}{\eta}$, leading to $e_1^S = \eta \sqrt{2p(1-q) (w_A(\underline{m}, r_G) - w_A(\underline{m}, r_B))}$ and $e_2^S = \frac{1-\eta}{\eta} e_1^S$. Using (2.4), this yields

$$U_S(e^S) = \left(\eta^2 + (1-\eta)^2 \right) \sqrt{2p(1-q) (w_A(\underline{m}, r_G) - w_A(\underline{m}, r_B))}$$

The supervisor prefers to induce e^S rather than e^* if $U_S(e^*) < U_S(e^S)$. Now consider an increase in $w_A(\underline{m}, r_B)$. This gives

$$\begin{aligned} \frac{\partial U_S(e^*)}{\partial w_A(\underline{m}, r_B)} &= -(\eta\kappa + (1-\eta)(1-\kappa)) \\ &\quad \times \frac{p(1-q)}{\sqrt{2(p(1-q) + q) w_A(\underline{m}, r_G) - 2p(1-q)w_A(\underline{m}, r_B)}} \\ \frac{\partial U_S(e^S)}{\partial w_A(\underline{m}, r_B)} &= -\left(\eta^2 + (1-\eta)^2 \right) \frac{p(1-q)}{\sqrt{2p(1-q) (w_A(\underline{m}, r_G) - w_A(\underline{m}, r_B))}} \end{aligned}$$

It follows that if $U_S(e^*) < U_S(e^S)$, we also have $\frac{\partial U_S(e^*)}{\partial w_A(\underline{m}, r_B)} < \frac{\partial U_S(e^S)}{\partial w_A(\underline{m}, r_B)}$. If the supervisor prefers some feasible e^S over e^* , an increase in $w_A(\underline{m}, r_B)$ does not induce the supervisor to demand e^* . Hence, given that the agent's rents increases in $w_A(\underline{m}, r_B)$ as shown above, it is optimal to set $w_A(\underline{m}, r_B) = 0$. After a bad report, the agent optimally receives no pay, regardless of measured performance.

■

Proof of Proposition 1. Condition (i) follows from Lemma 2, while verifiable performance pay $b(m = \underline{m}) = 0$ and $b(m \neq \underline{m}) = -c$ follows from Lemma 1. Let $\varphi \leq \frac{1}{2}$. Given c , \underline{m} , $b(m = \underline{m}) = 0$, and $b(m \neq \underline{m}) = -c$, the supervisor maximizes utility (2.4) with respect to requested effort levels \underline{e}_1 and \underline{e}_2 , subject to the agent's incentive compatibility constraint. If the supervisor requests effort levels that yield $m = \underline{m}$, this constraint is given by (2.21). If the first term between braces in (2.21) is larger than the second term, the supervisor optimally requests

$$\underline{e}_1 = \frac{\varphi}{\varphi^2 + (1 - \varphi)^2} \underline{m} + \frac{(1 - \varphi)}{\varphi^2 + (1 - \varphi)^2} \sqrt{2 \left(\varphi^2 + (1 - \varphi)^2 \right) (p + (1 - p) q) c - \underline{m}^2} \quad (\text{A6})$$

$$\underline{e}_2 = \frac{(1 - \varphi)}{\varphi^2 + (1 - \varphi)^2} \underline{m} - \frac{\varphi}{\varphi^2 + (1 - \varphi)^2} \sqrt{2 \left(\varphi^2 + (1 - \varphi)^2 \right) (p + (1 - p) q) c - \underline{m}^2} \quad (\text{A7})$$

Anticipating this, the principal maximizes utility (2.1) with respect to c and \underline{m} . The first-order conditions for c and \underline{m} are, respectively, given by

$$\begin{aligned} \frac{(1 - 2\varphi) (p + (1 - p) q)}{\sqrt{2 \left(\varphi^2 + (1 - \varphi)^2 \right) (p + (1 - p) q) c - \underline{m}^2}} - 1 &= 0 \\ \frac{1}{\varphi^2 + (1 - \varphi)^2} \left(1 - \frac{(1 - 2\varphi) \underline{m}}{\sqrt{2 \left(\varphi^2 + (1 - \varphi)^2 \right) (p + (1 - p) q) c - \underline{m}^2}} \right) &= 0 \end{aligned}$$

This can be solved for $c = \underline{m} = (p + (1 - p) q)$. Substituting for c and \underline{m} in (A6) and (A7) yields effort levels $\underline{e}_1 = \underline{e}_2 = (p + (1 - p) q)$.

This solution is not attainable if the agent prefers to feign compliance to the supervisor request by generating $m = \underline{m}$ at lower effort cost through effort levels along the performance measure's effort ratio (PMR). This arises if the second

term between braces in (2.21) is larger than the first term. Substituting for c and \underline{m} yields condition $p \geq \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$. The second condition on p follows from the supervisor's incentive to ignore performance target \underline{m} and request effort levels along her preferred effort ratio (SR), forcing the agent into manipulation of measured performance into \underline{m} when possible. When ignoring \underline{m} , the supervisor optimally requests

$$\underline{e}_1 = \eta \sqrt{\frac{2p(1-q)c}{\eta^2 + (1-\eta)^2}} \quad (\text{A8})$$

$$\underline{e}_2 = (1-\eta) \sqrt{\frac{2p(1-q)c}{\eta^2 + (1-\eta)^2}} \quad (\text{A9})$$

where we have used that the agent's incentive compatibility constraint is now given by

$$(1-q)c - \frac{1}{2}\underline{e}_1^2 - \frac{1}{2}\underline{e}_2^2 \geq (1-q)(1-p)c$$

Note that the agent cannot feign compliance in this case, as the principal denies the bonus after learning that $m \neq \underline{m}$. Comparing supervisor utility levels from adhering to and ignoring \underline{m} gives the following condition under which the supervisor prefers to adhere to performance target \underline{m} :

$$(p + (1-p)q) \geq \sqrt{2(\eta^2 + (1-\eta)^2)p(1-q)(p + (1-p)q)}$$

which can be rewritten to $p \leq \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$ as given in the proposition. ■

Proof of Proposition 2. By Lemma 2, it is not possible to induce $\frac{e_2}{e_1} > \frac{1-\varphi}{\varphi}$, while verifiable performance pay $b(m = \underline{m}) = 0$ and $b(m \neq \underline{m}) = -c$ follows from Lemma 1. For given c and \underline{m} , the supervisor maximizes utility (2.4) with respect to requested effort levels \underline{e}_1 and \underline{e}_2 , subject to the agent's incentive compatibility constraint (2.21). Assuming the first term between braces in (2.21) is larger than the second term, the supervisor optimally requests effort levels as given by (A6)

Appendix

and (A7). Substituting for \underline{e}_1 and \underline{e}_2 into the principal's utility (2.1) gives

$$U_P = \frac{1}{\varphi^2 + (1 - \varphi)^2} \underline{m} + \frac{(1 - 2\varphi)}{\varphi^2 + (1 - \varphi)^2} \sqrt{2 \left(\varphi^2 + (1 - \varphi)^2 \right) (p + (1 - p) q) c - \underline{m}^2} - a - c - w_s \quad (\text{A10})$$

By $\varphi > \frac{1}{2}$, the second term in (A10) is non-positive and decreasing in c . Note that the second term gives the principal's benefit of allowing the supervisor to implement an effort ratio different from (PMR). This implies that the optimal solution must have the smallest c possible given \underline{m} , as determined by

$$2 \left(\varphi^2 + (1 - \varphi)^2 \right) (p + (1 - p) q) c - \underline{m}^2 = 0 \quad (\text{A11})$$

Using this condition to substitute for c in (A10) and maximizing with respect to \underline{m} gives first order condition

$$\frac{1}{\varphi^2 + (1 - \varphi)^2} - \frac{\underline{m}}{\left(\varphi^2 + (1 - \varphi)^2 \right) (p + (1 - p) q)} = 0$$

which yields $\underline{m} = (p + (1 - p) q)$. Substituting for \underline{m} in (A11) gives $c = \frac{(p + (1 - p) q)}{2(\varphi^2 + (1 - \varphi)^2)}$, and the effort levels follow from substituting for \underline{m} and c in (A6) and (A7).

As the agent exerts effort levels along the PMR, the two terms between braces in (2.21) coincide, implying that feigning compliance by generating $m = \underline{m}$ with different effort levels never benefits the agent. The supervisor prefers adhering to the principal's demand $m = \underline{m}$ rather than deviating to the best-possible effort allocation along her most-preferred effort ratio (SR) when

$$(\eta\varphi + (1 - \eta)(1 - \varphi)) \sqrt{(p + (1 - p) q)} \geq \sqrt{\left(\eta^2 + (1 - \eta)^2 \right) \left(\varphi^2 + (1 - \varphi)^2 \right) p(1 - q)}$$

where the left-hand side (right-hand side) follows from substituting for the effort

levels (A6) and (A7) ((A8) and (A9)) into the supervisor's utility (2.4). Rewriting yields the condition on p as given in the proposition. ■

Proof of Proposition 3. If $p < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, the optimal contract derived in Proposition 1 is not attainable. The agent has an incentive to deviate to the effort levels on the performance measure's effort ratio (PMR) that generate $m = \underline{m}$ at lower effort cost (i.e. in the agent's incentive compatibility constraint (2.21), the second term in braces is larger than the first). Anticipating this, the supervisor would maximize her utility (2.4) with respect to effort requests \underline{e}_1 and \underline{e}_2 . Assuming (for now) that the supervisor requests satisfy performance target $m = \underline{m}$, the optimal effort request is given by

$$\begin{aligned}\underline{e}_1 &= \frac{\varphi}{\varphi^2 + (1-\varphi)^2} \underline{m} + \frac{(1-\varphi)}{\varphi^2 + (1-\varphi)^2} \sqrt{2(\varphi^2 + (1-\varphi)^2)pc} \\ \underline{e}_2 &= \frac{(1-\varphi)}{\varphi^2 + (1-\varphi)^2} \underline{m} - \frac{\varphi}{\varphi^2 + (1-\varphi)^2} \sqrt{2(\varphi^2 + (1-\varphi)^2)pc}\end{aligned}$$

implying that given $\varphi < \frac{1}{2}$, the contract specified in Proposition 1 (with $c = \underline{m}$) would lead to $\frac{\varphi}{1-\varphi} < \frac{\underline{e}_1}{\underline{e}_2} < 1$.

This outcome would yield higher rents to the agent than exerting zero effort, as determined by the first term between braces in (2.21). It follows that the principal can achieve the same effort allocation at lower cost by adjusting c and \underline{m} such that the two terms between braces in (2.21) are equal. In other words, the principal is constrained by the agent's temptation to resort to effort allocations on the PMR. This constraint is given by equating the two terms between braces in (2.21), which yields $c = \frac{1}{2(1-p)q} \frac{1}{\varphi^2 + (1-\varphi)^2} \underline{m}^2$. The optimal contract follows from substituting for effort levels (A6) and (A7) and for c into the principal's utility (2.1) and maximizing with respect to a , \underline{m} , and w_s (as before, verifiable performance pay $b(m = \underline{m}) = 0$ and $b(m \neq \underline{m}) = -c$ follows from Lemma 1). The first-order condition for \underline{m} is given by

$$1 + (1-2\varphi) \sqrt{\frac{p}{(1-p)q}} - \frac{1}{(1-p)q} \underline{m} = 0$$

Appendix

This can be rewritten to $\underline{m} = (1 - p)q + (1 - 2\varphi) \sqrt{p(1 - p)q}$, yielding $c = \frac{\left(\sqrt{(1-p)q} + (1-2\varphi)\sqrt{p}\right)^2}{2(\varphi^2 + (1-\varphi)^2)}$. The effort levels follow from substituting for c into (A6) and (A7). Given that $p < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, we have $c < p + (1 - p)q < \underline{m}$.

Lastly, we must show that the supervisor optimally adheres to $m = \underline{m}$. The best deviation is to request effort levels along his preferred effort ratio (SR), and manipulate measure performance into \underline{m} when possible. In that case, the agent receives no bonus if the principal observes true performance m , which happens with probability q . This implies that the agent's incentive constraint given deviation equals

$$(1 - q)c - \frac{1}{2}\underline{e}_1^2 - \frac{1}{2}\underline{e}_2^2 \geq (1 - q)(1 - p)c$$

Note that if the supervisor deviates, the agent has no incentive to feign compliance because as if $m \neq \underline{m}$, the bonus is forfeit if the agent cannot manipulate measured performance into \underline{m} . Given bonus c , the deviating supervisor optimally requests effort levels

$$\begin{aligned} \underline{e}_1 &= \eta \sqrt{\frac{2(1 - q)pc}{\eta^2 + (1 - \eta)^2}} \\ \underline{e}_2 &= (1 - \eta) \sqrt{\frac{2(1 - q)pc}{\eta^2 + (1 - \eta)^2}} \end{aligned} \tag{A12}$$

It is easily derived that given $\varphi < \frac{1}{2}$, the incentive to deviate is strongest for the supervisor most biased towards task 1, i.e. with $\eta = 1$. Below we show that even supervisors with $\eta = 1$ prefer to adhere to $m = \underline{m}$. Given $\eta = 1$, this implies that we have to show that effort in task 1 is lower when the supervisor deviates. Substituting for c , \underline{m} , and $\eta = 1$ into (A6) and (A12) implies that the supervisor

adheres to $m = \underline{m}$ when

$$\frac{1}{\varphi^2 + (1 - \varphi)^2} \left(\varphi(1 - p)q + (1 - 2\varphi^2) \sqrt{p(1 - p)q} + (1 - \varphi)(1 - 2\varphi)p \right) > \left(\sqrt{(1 - p)q} + (1 - 2\varphi) \sqrt{p} \right) \sqrt{\frac{(1 - q)p}{(\varphi^2 + (1 - \varphi)^2)}}$$

It can be shown that this expression increases in q . Rewriting condition $p < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, the lowest value of q considered in Proposition 3 is given by $q = \frac{p}{(1-p)(1-2\varphi)^2}$. Substituting for this level of q yields

$$\frac{2p(\varphi^2 + (1 - \varphi)^2)}{(1 - 2\varphi)^2} \left(\sqrt{\varphi^2 + (1 - \varphi)^2} - \sqrt{\left((1 - 2\varphi)^2 - \frac{p}{1 - p} \right)} \right) \geq 0$$

which holds for any φ and p , as $\varphi^2 + (1 - \varphi)^2 \geq (1 - 2\varphi)^2$ given that $0 \leq \varphi \leq 1$.

Hence, supervisors optimally adhere to $m = \underline{m}$. ■

Proof of Proposition 4. If $\varphi \leq \frac{1}{2}$ and $p > \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$, the optimal contract derived in Proposition 1 is not attainable. Similarly, if $\varphi > \frac{1}{2}$ and $p > \frac{(\eta\varphi + (1-\eta)(1-\varphi))^2}{(\eta-\varphi)^2} \frac{q}{(1-q)}$, the optimal contract derived in Proposition 2 is not feasible. In both cases, the supervisor has an incentive to deviate from inducing efforts that would satisfy $m = \underline{m}$ (i.e. (A6) and (A7)) to efforts along her most-preferred effort ratio (SR), as given by (A8) and (A9). Anticipating this, the principal must design a contract that meets the supervisor's incentive compatibility constraint:

$$\begin{aligned} \frac{\eta\varphi + (1 - \eta)(1 - \varphi)}{\varphi^2 + (1 - \varphi)^2} \underline{m} + \frac{\eta - \varphi}{\varphi^2 + (1 - \varphi)^2} \sqrt{2(\varphi^2 + (1 - \varphi)^2)(p + (1 - p)q)c - \underline{m}^2} \\ \geq \sqrt{2(\eta^2 + (1 - \eta)^2)(1 - q)pc} \end{aligned} \tag{A13}$$

where the left-hand side gives the supervisor's utility when meeting the principal's target and the right-hand side gives his utility when ignoring this target, both

following from substituting the effort levels into supervisor utility (2.4).

As before, verifiable performance pay $b(m = \underline{m}) = 0$ and $b(m \neq \underline{m}) = -c$ follows from Lemma 1. The optimal contract follows from substituting for effort levels (A6) and (A7) into the principal's utility (2.1) and maximizing with respect to c and \underline{m} , taking into account the limited liability constraints and the binding supervisor's incentive compatibility constraint (A13). Solving the Lagrangian gives lengthy first-order conditions, which, after straightforward (but tedious) rewriting, yield the expressions for c and \underline{m} given in the proposition. Substituting for c and \underline{m} into effort levels (A6) and (A7) yields the expressions for e_1 and e_2 .

■

Proof of Proposition 5. Substituting for $a = w_s = 0$ into the principal's utility (2.1) gives the principal's equilibrium payoff as $U_p = e_1 + e_2 - c$. Substituting for the optimal bonus c and equilibrium effort as given in Propositions 1-4 yields the following expressions for the principal's payoff U_p^i , where the superscript $i \in [1, 4]$ indicates the corresponding proposition (with some abuse of notation):

$$\begin{aligned} U_p^1 &= p + (1-p)q \\ U_p^2 &= \frac{p + (1-p)q}{2(\varphi^2 + (1-\varphi)^2)} \\ U_p^3 &= \frac{(\sqrt{(1-p)q} + (1-2\varphi)\sqrt{p})^2}{2(\varphi^2 + (1-\varphi)^2)} \\ U_p^4 &= \frac{(\sqrt{p(1-q)} + |1-2\eta|\sqrt{q})^2}{2(\eta^2 + (1-\eta)^2)} \end{aligned}$$

Note that all payoffs are identical to the corresponding levels of bonus pay c . Hence, all comparative statics are identical too.

First, we determine the comparative statics within each proposition.

(i) The (weakly) positive effect of p follows directly for Propositions 1, 2, and 4. For Proposition 3, we have $\frac{\partial U_p^3}{\partial p} = -\frac{\sqrt{pq} - (1-2\varphi)\sqrt{(1-p)q}}{2\sqrt{p}\sqrt{(1-p)q}} > 0$, where the

sign follows from the conditions $\varphi \leq \frac{1}{2}$ and $p < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$ under which this Proposition is relevant.

(ii) The (weakly) positive effect of q follows directly for Propositions 1, 2, and 3. For Proposition 4, we have $\frac{\partial U_p^4}{\partial q} = -\frac{\sqrt{q}p-(1-2\eta)\sqrt{p(1-q)}}{2\sqrt{q}\sqrt{p(1-q)}} > 0$, where the sign follows from the conditions $p > \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$ or from $\eta > \varphi > \frac{1}{2}$ and $p > \frac{(\eta\varphi+(1-\eta)(1-\varphi))^2}{(\eta-\varphi)^2} \frac{q}{(1-q)}$ under which this Proposition is relevant.

(iii) Under the conditions governing Propositions 1 and 4, φ has no effect on U_p . In Proposition 2, the principal's payoff is decreasing in the bias $|\varphi - \frac{1}{2}|$. In Proposition 3, we have $\frac{\partial U_p^3}{\partial \varphi} = -\frac{(1-2\varphi)(p-q+pq)+4\varphi(1-\varphi)\sqrt{p(1-p)q}}{(\varphi^2+(1-\varphi)^2)^2}$, which is also decreasing in bias $|\varphi - \frac{1}{2}|$ when $p < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$.

(iv) The supervisor's bias matters only under the conditions that make Proposition 4 relevant. Then, $\frac{\partial U_p^4}{\partial \eta} = -\frac{(1-2\eta)(q-p+pq)+4\eta(1-\eta)\sqrt{p(1-q)q}}{(\eta^2+(1-\eta)^2)^2}$, which decreases in supervisor bias $|\eta - \frac{1}{2}|$ when $p > \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$ as well as when $p > \frac{(\eta\varphi+(1-\eta)(1-\varphi))^2}{(\eta-\varphi)^2} \frac{q}{(1-q)}$ and $\eta > \varphi > \frac{1}{2}$.

Second, we compare the principal's payoffs at the exact parameter thresholds that determine which proposition is relevant. Substituting for $\varphi = \frac{1}{2}$ into U_p^2 , we have $U_p^1 = U_p^2$. Similarly, when $p = \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, we have $U_p^1 = U_p^3$, and when $p = \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$ we have that $U_p^1 = U_p^4$. Lastly, when $p = \frac{(\eta\varphi+(1-\eta)(1-\varphi))^2}{(\eta-\varphi)^2} \frac{q}{(1-q)}$ and $\eta > \varphi > \frac{1}{2}$ we have that $U_p^2 = U_p^4$. Hence, given $\eta > \varphi$, marginal changes in parameter values do not lead to jumps in the principal's payoff. ■

Proof of Proposition 6. We focus again on the case of $\eta > \varphi$. Comparing the principal's payoff when using objective performance evaluation only and subjective performance evaluation only, as given by (2.11) and (2.19), respectively, to the payoffs when combining them, as given by U_p^i as defined in the proof to Proposition 4, it follows directly that U_p^1 and U_p^2 are (weakly) larger than (2.11) and (2.19). Hence, if both p and q are large enough (i.e. when $\frac{1}{(1-2\eta)^2} \frac{q}{(1-q)} \geq p \geq \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$), combining objective and subjective performance evaluation yields lower rents to the agent and, unless $\eta > \varphi > \frac{1}{2}$, eliminates the bias that might arise when using only objective or only subjective perfor-

mance evaluation. Both effects increase the principal's payoff. Next, suppose that $p < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, such that the outcome with combined evaluation is biased towards PMR (Proposition 3). Even when the supervisor would be unbiased ($\eta = \frac{1}{2}$), the principal benefits from also using the verifiable performance measure: U_p^3 exceeds (2.19) for any $p < \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$. Similarly, if $p > \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$, such that combined evaluation results in a bias towards the SR, the principal benefits from also using subjective performance evaluation, as U_p^4 is larger than (2.11) for any $p > \frac{1}{(1-2\eta)^2} \frac{q}{(1-q)}$ even when $\varphi = \frac{1}{2}$. ■

Proof of Proposition 7. If $p_S < \frac{(\eta_S\varphi+(1-\eta_S)(1-\varphi))^2}{(\eta_S-\varphi)^2} \frac{q}{(1-q)}$ and $\varphi \leq \frac{1}{2}$ or $\varphi > \eta_S > \frac{1}{2}$, the principal's payoff with both supervisors is either U_p^1 or U_p^3 , which are both increasing in p and independent of η . Hence, the principal prefers the specialist. Similarly, if $p_S > \frac{1}{(1-2\eta_S)^2} \frac{q}{(1-q)} > p_G$, the specialist yields principal's payoff U_p^4 , while the generalist yields payoff U_p^1 . Result (i) in Proposition 5 showed that starting from U_p^1 , an increase in p such that the principal is forced to allow for a bias, yielding payoff U_p^4 , benefits the principal for any η . This proves the first part. When $\eta_S > \varphi > \frac{1}{2}$ and $p_S < \frac{(\eta_S\varphi+(1-\eta_S)(1-\varphi))^2}{(\eta_S-\varphi)^2} \frac{q}{(1-q)}$, employing the specialist supervisor yields payoff U_p^2 . Employing the generalist yields U_p^1 if $p_G \geq \frac{q(1-2\varphi)^2}{1+q(1-2\varphi)^2}$, otherwise it yields $U_p = \frac{(\sqrt{(1-p)q}+(2\varphi-1)\sqrt{p})^2}{2(\varphi^2+(1-\varphi)^2)}$. Note that the latter expression mirrors U_p^3 , and holds for $\eta < \varphi$ and $\varphi > \frac{1}{2}$ (here relevant when employing a generalist as $\eta_G = \frac{1}{2}$), while U_p^3 is derived under the conditions $\eta < \varphi$ and $\varphi > \frac{1}{2}$ (see Proposition 3). Comparing these payoffs yield the two conditions in case (i), respectively. Lastly, when $p_G > \frac{1}{(1-2\eta_S)^2} \frac{q}{(1-q)}$, the specialist yields principal's payoff U_p^4 , while the generalist yields payoff U_p^1 . Comparing these payoffs gives the condition in the proposition. ■

Chapter 3

Team Incentives, Task Assignment, and Performance: A Field Experiment

Joint with Robert Dur and Josse Delfgaauw¹

3.1 Introduction

In many organisations, employees work in teams that perform a variety of tasks. A team's ultimate performance depends on how well employees perform their tasks as well as on the division of tasks among the employees. For instance, allocating more important tasks to more talented employees will often improve a team's performance. However, in practice, performance considerations are rarely the only determinant of task allocation, in particular when some tasks are more

¹This chapter is forthcoming in *The Leadership Quarterly* (2018).

interesting or pleasant than others. In such cases, preferences of team members for tasks may affect the allocation, as well as fairness concerns or seniority. When team managers decide on task allocation, favouritism may play a role as well. The importance of these factors can be diminished by introducing or strengthening incentive pay based on team performance for teams and their managers. In addition to inducing workers to perform better on their tasks, team incentives may induce teams or their managers to reallocate tasks in a performance-enhancing way.

In this paper, we present the results of a field experiment designed to study the effects of team incentive pay on team performance and task allocation in teams. The experiment took place in a retail chain in The Netherlands comprising 108 geographically dispersed stores. We randomly selected 60 stores to participate in a short-term sales tournament. Each participating sales team competed with two comparable sales teams for a period of six weeks on the basis of sales relative to a pre-determined sales target. Employees and the manager of the best-performing store earned a bonus of 50 euro each, which amounts to more than 3% of monthly employee earnings. During the tournament, participating stores received weekly feedback informing them about the current ranking in their group.

We use administrative sales data to analyse the effects of the team incentive on performance. Furthermore, we conducted surveys among employees and managers of all stores before and after the tournament period to learn about task allocation in stores. The surveys ask store employees and managers about the importance of several aspects for task allocation in their team, including employee ability, employee preferences over tasks, fairness concerns, seniority, and managerial favouritism. In addition, we collected data on employees' job satisfaction. By conducting identical surveys before and after the tournament in both the treatment and control group, we are able to estimate the effect of the team incentive on task allocation within teams (as perceived by employees) and on job satisfaction.²

²An alternative measure of task allocation would be to ask employees about the tasks they

Introduction

The theoretical predictions are as follows. The team incentive increases the importance of team performance to employees and managers. Consequently, they should have a stronger incentive to exert effort, leading to better performance. Furthermore, team performance should play a more important role in the allocation of tasks among employees. Hence, we predict that employee ability becomes more important in task assignment, while the other considerations (employee preferences, fairness concerns, seniority, and favouritism) become less important. This revised task allocation should also result in better team performance. Job satisfaction may suffer as workers' individual task preferences play a smaller role in task assignment. On the other hand, some workers' job satisfaction may actually increase as considerations like favoritism and seniority are muted. The overall implications for job satisfaction are thus theoretically unclear.

Our experiment follows the tradition in organizational economics to focus on incentives and contract design. However, by hypothesizing about and collecting detailed data on how leaders assign tasks to employees, we move a step in the direction of the literature on leadership in management and psychology, which allows for a much richer role of leadership than the economics literature (Zehnder et al. 2017). In particular, we stress the role of leaders in coordinating employees' complementary actions that jointly determine the team's success. We share this feature with Burgess et al. (2010), who analyse the introduction of team pay-for-performance at the UK tax authorities. The incentive scheme covered only a part of the tasks that teams are responsible for. The findings indicate that team incentives increased performance, part of which can be attributed to a change in task assignment within teams, as managers disproportionately reallocated efficient workers to the incentivised tasks. Two other papers in economics studying task assignment by managers are Bandiera et al. (2007 and 2009). They introduce incentive pay for managers in a UK fruit farm, making their pay dependent of their subordinates' performance. This induced managers to assign more

have actually performed. Any observed changes would be hard to interpret, though, because we lack reliable information about the relation between task allocation and performance.

productive employees to the incentivised task (picking fruit), leading to a substantial increase in productivity. By contrast, when paid a flat wage, managers were more likely to assign this task to employees who were socially connected to them. Hence, providing performance-based incentives to managers reduced favouritism in task assignment. Contrary to these earlier studies, we examine the effects of an incentive scheme that rewards overall team performance, not performance on a subset of tasks. We lack administrative data on individual task assignment and productivity. Instead, we use surveys to assess whether the incentives affect how tasks are allocated, as perceived by the employees.³

Our results are as follows. We find no effect of the team incentive on sales performance. This average treatment effect is fairly precisely estimated. Furthermore, we find no evidence supporting the hypothesis that task allocation has changed to enhance performance in treated stores. In particular, we find no effect on the importance of employee ability in the allocation of tasks. Nor do employees in treated stores report more often that "the division of tasks is such that the best possible sales performance is achieved". The average treatment effect on job satisfaction is not statistically significant either. Together, these results suggest that business continued as usual despite the treatment. We provide an extensive discussion of possible interpretations of the null results in the final section of our paper.

The estimated average treatment effect in this study is at the lower end of the range of estimates in comparable studies in the retail sector. In Delfgaauw et al. (2013, 2014, 2015), we implement comparable team-based tournament incentives schemes across shops of different retail chains and obtain average treatments effects on performance ranging from 0% to 5% increase in sales. Friebe et al. (2017) implement a (non-competitive) team incentive based on performance targets in a German bakery chain and find an average treatment effect of 3%.

³Another recent paper examines task assignment in teams in the lab. Cooper and Sutter (2018) compare teams with random task assignment to teams with endogenous task assignment, finding that the latter do not outperform the former, despite a positive selection effect. In contrast to our setting, their teams have no leader. Muehlheusser et al. (2016) study task allocation by managers in the context of professional soccer.

Introduction

Casas-Arce and Martinez-Jerez (2009) find a more substantial effect of a tournament incentive implemented among independent retailers of a commodities firm in which there was substantially more money at stake than in the current study. We deviate from these previous studies in our focus on task allocation within teams. By contrast, earlier studies have looked at how treatment effects relate to the gender composition of the team and gender of the manager (Delfgaauw et al. 2013), the prospect of participation in further tournament rounds and volatility in performance (Delfgaauw et al. 2015), the extent to which employees are able to influence waiting times (Friebel et al. 2017), and intermediate rankings during the tournament (Casas-Arce and Martinez-Jerez 2009, Delfgaauw et al. 2014). The emerging evidence suggests that even within a sector, comparable incentive schemes can lead to different responses across organisations.

Experiments on the use of team incentives have also been conducted in other settings. Erev et al. (1993) recruit students to pick oranges and find that participants grouped into teams of 4 under team-based pay are 30% less productive than participants under individual incentive pay. This negative effect of team-based pay is mitigated when the teams competed under a tournament incentive scheme. Bandiera et al. (2013) study endogenous team formation and find that providing relative performance information reduces performance while relative performance pay increases performance, partially due to changes in team composition. Studies in education have found mixed effects of introducing team incentive pay for teachers (Lavy 2002, Glewwe et al. 2010, Springer et al. 2010, Muralidharan and Sundararaman 2011, Goodman and Turner 2012, Fryer 2013).⁴

The relationship between performance-related pay and job satisfaction is the subject of a small literature. With a single exception (Friebel et al. 2017), the existing evidence is correlational. Heywood and Wei (2006) and Green and Heywood (2008) document a positive relation between job satisfaction and per-

⁴In a lab experiment, Chen and Lim (2013) find that contests between teams of participants yield higher productivity than contests between individuals, but only when (potential) teammates could meet before the contest. Babcock et al. (2015) offer students a reward for meeting a target regarding study room visits and find that team-based pay can outperform individual-based pay, but only when the teammates know each other.

formance pay, including profit-sharing, in US and British panel survey data, respectively. Studying the introduction of incentive pay in two firms, Welbournel and Cable (1995) document a similar relation. Using cross-sectional British data, Petrescu and Simmons (2008) find no relation between team performance pay and job satisfaction. For individual incentive pay, they find a positive relation with satisfaction with pay, but no relation with overall job satisfaction. McCausland et al. (2005) find a positive relation between incentive pay and job satisfaction for highly paid workers but the reverse for lower paid workers. We add to this literature by examining the effect of introducing a short-term team incentive on job satisfaction using an experimental design, allowing for a causal interpretation of our findings.

The rest of this paper is organized as follows. The next section discusses the experimental setting and design. In Section 3.3 we discuss the estimation procedure. Descriptive statistics are presented in Section 3.4 and the estimation results in Section 3.5. Section 3.6 discusses our findings and concludes.

3.2 Experimental Context and Design

3.2.1 Experimental context

The experiment took place from October 2013 to January 2014 among 108 stores of a retail chain selling lingerie and swimwear in the Netherlands. All stores are company-owned, there are no franchisers. All managers and employees are female. A store has a manager and on average 7 employees. The majority of employees works part-time or on-call.

Employees earn an hourly wage slightly above the legal minimum hourly wage. They can occasionally earn team bonuses during incentive periods that are usually timed to coincide with marketing efforts. These bonuses are generally earmarked for team outings.⁵ The company's management was interested in conducting this

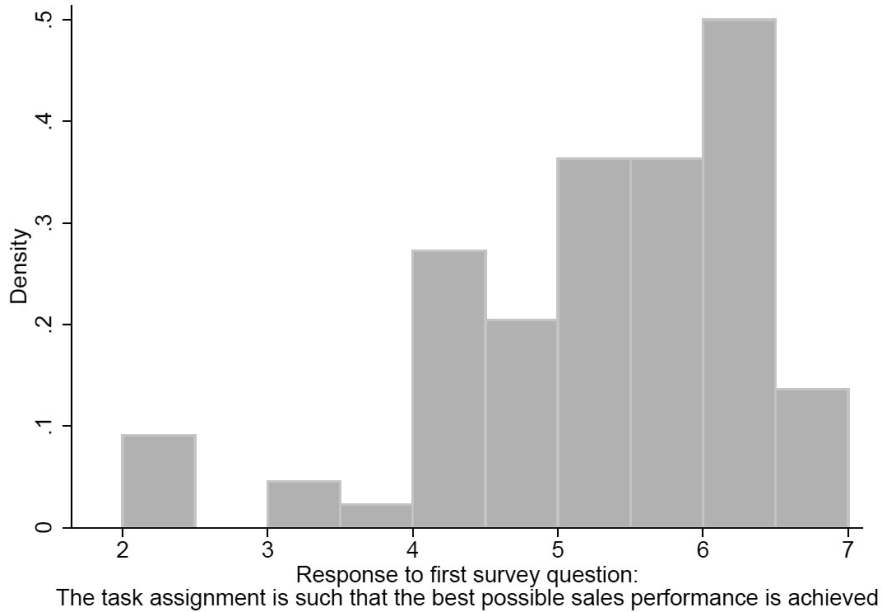
⁵No such incentive period ran concurrently with our experiment.

field experiment as it wished to explore a more extensive use of incentive pay.

Decisions regarding the product range, pricing, and marketing are made by the retail chain’s management. The primary tasks of store staff include advising customers, attending the register, administration, and keeping the displays stocked and tidy. According to the company’s management, employees have substantial influence on store performance, especially through an assertive and commercial attitude when advising customers. Furthermore, within the company it is widely acknowledged that store employees differ substantially in their ability to generate sales when advising customers. The chain’s management and the store managers we spoke to considered task allocation to be an important channel through which store managers affect store performance. This anecdotal evidence suggests that, in this retail chain, task allocation matters for team performance, providing a good setting for an experiment on whether team incentives affect task assignment.

The store manager is responsible for staffing of the store and has the authority to assign tasks to employees. In practice, employees are consulted and can express their preferences. Furthermore, employees typically perform multiple tasks, prioritizing advising customers and attending the register when there are many customers and cleaning and stocking during quiet moments. In the survey we conducted before the experiment, the average response of store managers to the question “I determine the allocation of tasks” equals 6.3 on a 7-point scale ranging from completely disagree (1) to neutral (4) to completely agree (7). The average response to the question “I monitor whether everyone performs their tasks well” equals 6.4 on the same 7-point scale. Hence, store managers feel that they are in charge of task allocation. The mean response among employees to the statement ‘I decide myself which tasks to perform’ was 4.3 and their mean response to the statement ‘The store manager decides about the allocation of tasks in the store’ was 5.2, both on a 7-point Likert scale. This suggests that the store manager coordinates the allocation of tasks, but that employees do have some leeway in deciding which tasks to perform at a given time. Importantly,

Figure 3.1: Task assignment and efficiency, store averages



we also asked employees whether the task allocation in their store achieved the best possible sales performance. The mean response to this statement was 5.2. Taking averages across employees at the store level, Figure 3.1 gives the distribution of responses across stores. This indicates that while task allocation is catered towards enhancing performance, employees in many stores do see room for improving sales performance through changes in task allocation. We discuss the surveys in more detail in subsection 3.2.3.

3.2.2 Experimental Design

We implemented tournaments between subsets of stores over a period of six weeks. A randomly selected subset of stores was assigned to groups of three stores each; we discuss the assignment procedure at length in subsection 3.2.4 below. Within these groups, stores competed for a prize. The performance measure in the tournament was a store's cumulative realized sales over the period of six weeks as

a percentage of cumulative sales targets. These weekly targets for each store’s sales are set by the company’s management at the start of the financial year, long before we set up the experiment.⁶ These targets take into account variation in sales due to e.g. seasonal effects, holidays, and planned marketing efforts, as well as store-specific factors. In our data, time and store-fixed effects account for 92.7% of the variation in sales targets. Realized sales relative to sales targets is a familiar performance metric to employees. Store managers receive this performance measure on a weekly basis. Let $r_{s,w}$ denote sales revenues realized by store s in week w , and $b_{s,w}$ the sales target for store s in week w . Cumulative performance of store s after w weeks is given by

$$P_{s,w} = \frac{\sum_1^w r_{s,w}}{\sum_1^w b_{s,w}} \cdot 100\% \quad (3.1)$$

All employees and the manager of the store with the best performance in the group after six weeks received a bonus. Full-time employees received €50, part-time employees received a bonus proportional to their contract size. The bonus amounts to approximately 2.5% of employees’ earnings in a six-week period. Depending on the size of the store and experience, store managers earn about 20 to 50 percent more than employees. Hence, the bonus amounts to 1.7% to 2.1% of store managers’ pay in a six-week period. This is comparable to Burgess et al. (2010), where employees could earn a bonus of about 3% of pay and managers a bonus of either 2% or 4%, depending on the treatment. In Bandiera et al. (2007, 2009), the bonus was considerably higher. It could reach up to 25% of total compensation, and actual bonus payout was about 7% of total pay.

Out of fairness considerations, the company’s management insisted on allowing each store the opportunity to compete. Therefore, we implement tournaments in two periods. The second tournament period started three weeks after the end of the first tournament period. In the first period, 60 randomly chosen stores

⁶Our experiment ran at the end of the financial year, and the sales targets had been determined well before the experiment was planned. Hence, the targets are not affected by the experiment.

participate while the remaining 48 stores comprised the control group. The relatively large number of stores in the treatment group serves two purposes. First, as we expect the variance of performance to be larger among treatment stores than among control stores during the tournament, a larger treatment group serves to increase the power of our analysis (List et al. 2011). Second, a larger treatment group enables us to have a set of stores compete in both periods, to analyse potential spill-over effects. In the second tournament period, all 48 stores that did not participate in the first period were assigned to participate. In addition, we randomly selected 15 stores that did participate in the first tournament period to also participate in the second period.

All communication regarding the tournaments was sent through regular company channels using company material. Store personnel was unaware of our involvement in the incentive scheme. Prior to the first tournament period, the company's management announced that several incentive events would be held in the near future. Stores were informed that – while each store would participate at least once – they would not necessarily participate in all events. Stores were not informed about an upcoming tournament if they would not be participating.⁷ In the communication towards the stores, there was no mentioning of a specific interest in task allocation.

The tournaments were announced and explained to participating stores in the week prior to the start of each tournament. The company decided to run the tournaments under the name 'Sexy Super Cup'.⁸ Participating stores received a large poster specifically designed for this event, which contained the rules of the contest and was supposed to be glued to a wall or door in the backoffice. During the tournament, each week the stores received a small poster with the ranking in their group (see Figure 3.2 for an example; the original poster was in Dutch), which could be glued to dedicated spaces on the large poster. This allowed stores to track their (relative) performance during the tournament.

⁷However, personnel may in some instances have learned of an ongoing tournament due to contacts with other stores.

⁸Pun intended.

3.2.3 Surveys

We have asked all store employees twice to complete an online survey. The first survey was administered prior to the first tournament period and the second survey after the first tournament period. The two surveys were identical. The goal of the surveys was to measure the importance of different considerations driving the allocation of tasks in the store, as well as employees' job satisfaction. Invitations to participate in the survey were sent on behalf of Erasmus University Rotterdam to all employees and managers of all 108 stores. The invitation for the first survey was sent three weeks before the start of the first tournament period, and the invitation for the second survey was sent one week after the end of the first tournament period. The invitations included a personal code that allowed us to link survey responses. Employees were given two weeks to complete each survey. In the second week we called stores as a reminder, using a call script. Neither the surveys nor the call script mentioned the tournaments. As an incentive to complete the surveys, one randomly chosen respondent in each wave was awarded with a tablet with a retail value of 150 euro.

In the surveys, respondents had to evaluate statements using a 7-point Likert scale. In particular, we asked employees about the importance of employee ability, employee preferences, fairness concerns, favouritism, and seniority in the determination of their store's task allocation in the past two months. Furthermore, we asked whether the store's task assignment is such that the best possible sales performance is achieved. These questions are presented in Table 3.1.

Conducting the survey before and after the first tournament period among both first-period treatment and first-period control stores allows us to analyse the effects of the tournament treatment on task allocation (as perceived by employees) and job satisfaction. We did not conduct a third wave of surveys after the second tournament period, as at that point all stores would have been treated at least once. This makes it impossible to distinguish between time-specific determinants of task allocation and changes in task allocation as a result of the treatment.

Table 3.1: Survey questions on task assignment

Variable	Survey question
Task assignment	Please indicate which answer best describes the situation in the past two months. The following statements are concerned with the task assignment in your store.
Ability	The division of tasks is such that everyone does that which she does best.
Preference	The division of tasks is such that everyone does that which she enjoys most.
Seniority	Employees who have been with the store longer carry out more pleasant tasks.
Fairness	The divisions of tasks is fair.
Favouritism	Friends of the manager carry out more pleasant tasks.
Efficiency	The division of tasks is such that the best possible sales performance is achieved.
Job satisfaction	How satisfied are you with your job at ...
The statements on task assignment and efficiency came with a 7-point Likert scale ranging from 'Completely disagree' to 'Completely agree'. The question on job satisfaction came with a 7-point Likert scale ranging from 'Very dissatisfied' to 'Very satisfied'.	

Figure 3.2: Example of intermediate ranking provided to stores



Store managers and store employees received similar surveys, where the wording in the survey for managers was slightly altered to reflect their specific role in allocating tasks.⁹ Managers may have stronger strategic or image concerns in answering survey questions regarding task allocation than employees. Therefore, in the analysis we only use the survey responses of employees.¹⁰ Response rates among employees were 34.5% (258 out of 747) on the first survey and 18.5% (140 out of 739) on the second survey. At the store level, we have at least one respondent for 82.4% of stores (89 out of 108) in the first survey and for 61.1% (66 out of 108) in the second survey. We will discuss possible selection bias in Section 3.4.

⁹In the survey for managers, we left out one item, the one on favouritism.

¹⁰Indeed, in the first survey, managers' average response to the question on the efficiency of task allocation is equal to 6.2 out of 7. This is significantly higher than employees' average response of 5.2 to this question.

3.2.4 Assignment procedure

For the first tournament period, we used a stratified random assignment procedure to create balanced treatment and control groups in terms of task allocation and prior performance, as follows. First, for each of the 89 stores with at least one respondent in the first survey, we calculated the average response to the statement ‘the division of tasks is such that everyone does that which she does best’, measuring the importance of employee ability in allocating tasks. We ranked these stores based on this score. We added one randomly chosen store from the set of 19 stores with no survey response, at a randomly chosen rank. From this ranking of 90 stores, we created 5 strata of 18 stores where the top 18 stores constituted one stratum, as well as stores ranked 19 to 36, and so on. The remaining 18 stores with no survey response constituted a sixth stratum. Within each of these six strata, we ranked stores based on average weekly performance (sales relative to sales target) over the 36 weeks prior to the experiment. We divided each stratum into two substrata constituting the top 9 and bottom 9 stores in terms of prior performance. Finally, we randomly selected five stores out of the 9 stores in each of the 12 substrata to participate in the first tournament period. Hence, the treatment group comprises 60 stores and the control group contains 48 stores.

In the second tournament period, all stores in the first-period control group participated in the tournament because the companies’ management wanted all stores to participate in at least one tournament. In addition, we selected 15 stores from the first-period treatment group. We randomly selected one first-period treatment store from each of the 12 substrata created in the first-period assignment procedure and added 3 more stores randomly chosen out of the remaining first-period treatment stores. The remaining first-period treatment stores did not participate in the second period competition.

In assigning the participating stores to groups of three stores in the tournament, we aimed at maximising the level of competition by grouping the stores

together on the basis of past performance. For the first tournament period, we ranked the participating stores by cumulative sales over 36 weeks prior to the experiment. Next, we grouped the three best performing stores together, as well numbers 4 to 6, and so on. To reduce possible sabotage opportunities, we made in total 4 adjustments to prevent that stores from the same regional area (Dutch province) were grouped together. For the second tournament period, we followed a similar procedure. We ranked the participating stores based on cumulative past performance over 45 weeks prior to the tournament period and divided them into groups of three similarly performing stores. This time, we had to make 8 adjustments to prevent stores within the same regional area from competing with one another. Upon informing the companies' management of the assignment, we learned that two stores, which had not participated in the first tournament period, would be closed for refurbishment during the second tournament period. We drop these stores' observations during and after their refurbishment from the analysis. We did not replace these two stores in the tournament, so that in two second-period tournament groups only two stores competed.¹¹ Hence, 61 stores participated in the second tournament period.

3.3 Estimation

We estimate the effect of participating in the experiment on weekly performance, task allocation, and job satisfaction using OLS with period-fixed effects and either store- or individual-fixed effects. The average treatment effect on stores' performance is estimated by

$$Y_{s,w} = \alpha_s + \gamma_w + \beta T_{s,w} + \varepsilon_{s,w}, \quad (3.2)$$

where $Y_{s,w}$ is weekly performance of store s in week w , measured as actual sales revenue over targeted sales: $Y_{s,w} = \frac{r_{s,w}}{b_{s,w}}$. Store and week-fixed effects are given by

¹¹We keep these four stores in the analysis below. None of the results change if we drop these stores after the first tournament period.

α_s and γ_w , respectively. $T_{s,w}$ is a dummy denoting whether store s is participating in a tournament in week w , so that β measures the average treatment effect. The error term $\varepsilon_{s,w}$ is clustered at the store level to account for possible serial correlation.¹² We will also estimate the effects separately for the first and second tournament period.

For the first tournament period, stores have been randomly allocated to either treatment or control, implying that in expectation the control group provides a reliable counterfactual for the performance of the treatment stores in the tournament. In the second tournament period, however, all non-participating stores have participated in the first period. If participation leads to carry-over effects, for instance due to learning or fatigue, these stores do not constitute a proper control group for the stores that compete for the first time in the second period. We use the 15 stores that participate in both waves to analyse possible carry-over effects. To prevent short-term carry-over effects from affecting the estimates, the three weeks in between the two tournament periods are excluded from the analysis.¹³

Estimation of the incentive effect on job satisfaction and task allocation is based on the first tournament period, as the surveys were administered before and after this period. We estimate the effect of the team incentive both at the individual employee level and at the store level. For the analysis at the store level, we use the average response on the survey items across all respondents employed in a given store. The average treatment effect is estimated by

$$R_{i,t} = \alpha_i + \gamma_t + \beta T_{i,t} + \varepsilon_{i,t}, \quad (3.3)$$

where $R_{i,t}$ is the survey response of unit i (individual or store) in survey t (before or after the first tournament period). Observation unit-fixed effects and survey-

¹²For ease of interpretation, we estimate the incentive effect on weekly performance as opposed to cumulative performance (as given by (3.1)). This is inconsequential for the estimation results since (3.1) determines performance by comparing total sales to total targeted sales over the tournament period. Thus sales staff cannot strategically focus efforts on apparently "easy" weeks as a sale counts equally towards performance regardless of the week in which it occurs.

¹³Including these weeks in the analysis does not influence the results.

Descriptive statistics

fixed effects are given by α_i and γ_t , respectively. $T_{i,t}$ is a dummy that takes value 1 for responses on the survey after the first-period tournament if the store (of individual) i was part of the first-period treatment group. Hence, β measures the average treatment effect. Finally $\varepsilon_{i,t}$ is the error term, clustered at the store level.

The store level is the natural level of analysis, as we randomized at the store level. Furthermore, task assignment affects the team as a whole. However, there are three caveats in analyzing the results at the store level. First, for a given store the respondents to the first survey may differ from the respondents to the second survey. Insofar as selection into and out of the survey is correlated with assignment to the first-period treatment group, this may bias the results. Below, we analyse the self-selection of employees into answering the first and second survey and find no indication of selection related to first-period assignment. Second, stores are given the same weight in the analysis independent of the number or fraction of employees that answered a given survey. Weighing stores by the number of respondents (in either the first or the second survey) does not affect the estimates. Third, store-level averages mask within-store differences in responses. Among the questions on task allocation in the first survey, the fraction of total variation explained by store-fixed effects ranges from 0.34 to 0.44. Hence, there is sizable heterogeneity across stores, but also considerable differences within stores. Therefore, we also present the average treatment effects estimated at the individual employee level, accounting for individual-fixed effects.

3.4 Descriptive statistics

We have weekly data on store performance covering a year starting in February 2013. In addition, prior to the first tournament period we received personnel data of all stores. This includes information on employees' age, tenure, contractual hours (measure in full-time equivalent, fte), and position. Figure 3.3 depicts weekly targeted and actual sales averaged across stores for the 52 weeks in our

dataset. Sales is highly volatile, but most of it is predicted by the company's management as sales and targeted sales follow by and large the same pattern.

Figure 3.3: Average sales and average targeted sales

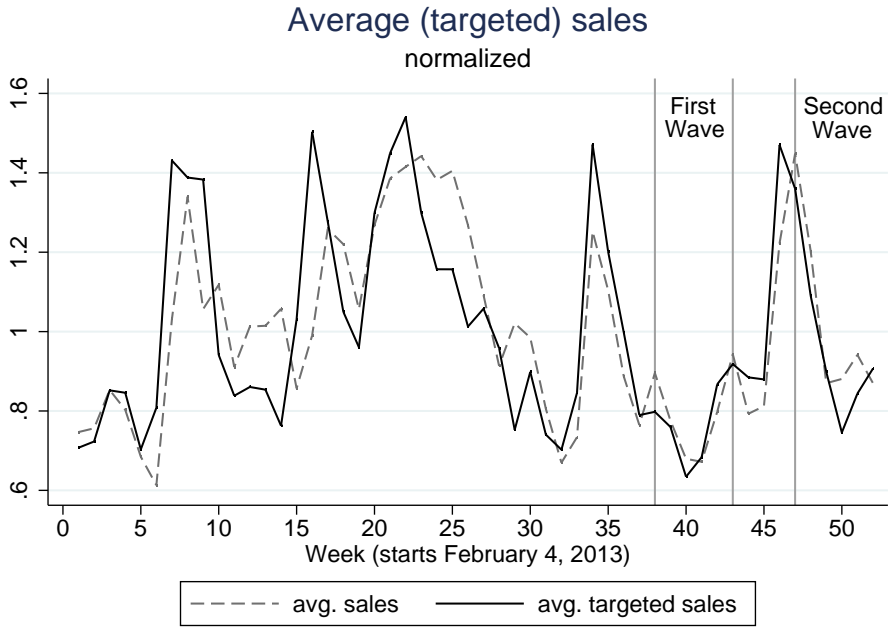


Table 3.2: Store characteristics, by first-period assignment

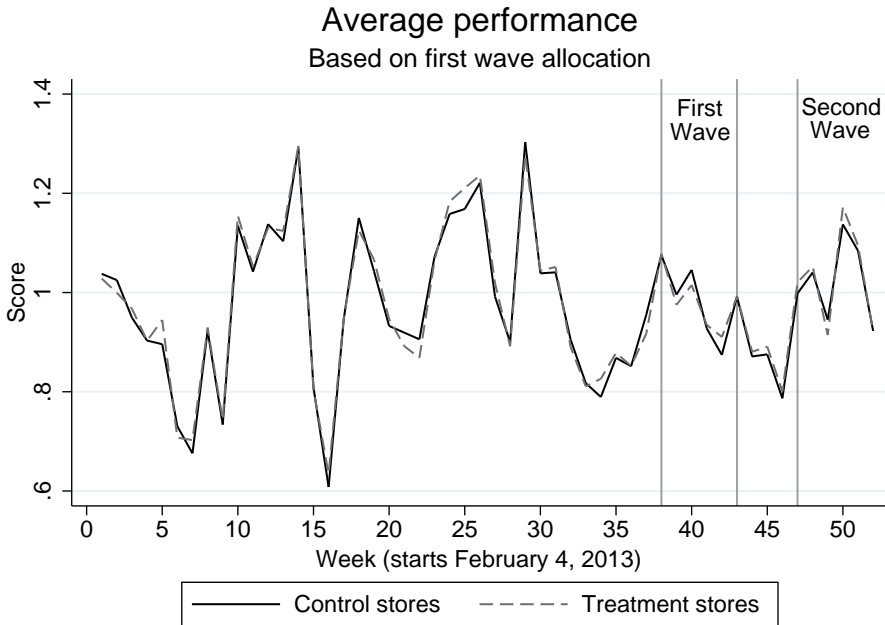
	Total		Control		Treatment	
	mean	sd	mean	sd	mean	sd
Average prior performance	0.95	0.08	0.95	0.10	0.96	0.07
Team size	7.91	3.30	8.19	3.61	7.68	3.04
Aides	1.91	1.56	1.75	1.62	2.03	1.51
Age manager	43.79	11.06	43.53	11.70	44.00	10.62
Tenure manager	15.43	13.88	13.86	8.94	16.67	16.80
Fte manager	0.89	0.08	0.89	0.07	0.88	0.09
Average age staff	38.40	7.06	38.33	6.77	38.46	7.34
Average tenure staff	8.70	4.94	9.05	5.34	8.42	4.61
Average fte staff**	0.48	0.13	0.51	0.13	0.45	0.12
Observations	108		48		60	

*, **, and *** indicate that the difference between treatment stores and control stores is statistically significant at the $p < .1$ level, $p < .05$ level, and $p < .01$ level, respectively.

Descriptive statistics

Descriptive statistics are given in Table 3.2. The first column shows that, on average, stores' sales fell short of sales targets by about 5 percent. Stores employ on average 8 individuals (including the manager), of whom two work on-call. Most regular employees work part-time. All employees and managers are female. Table 3.2 also reports these statistics separately for the treatment group and the control group in the first tournament period. This shows that the two groups are similar in terms of past performance and personnel characteristics. The only exception is average contract size ('fte') which is slightly but significantly larger in the control stores. Figure 3.4 depicts average performance separated by first-period assignment. This shows that both prior to and during the tournament periods performance in both groups was similar.

Figure 3.4: Performance of treatment and control stores.



Besides the performance and personnel data, we also have survey data. The response rates of the two surveys were 34.5% (258 out of 747) and 18.5% (140 out

of 739) for employees on the first and second survey respectively. 92 employees completed both surveys. We averaged employees' survey responses by store. This results in 89 and 66 stores with staff survey data for the first and second survey, respectively, and 60 stores for which we have respondents for both surveys. Due to some item non-response, the exact number of stores with at least one response varies a bit across questionnaire items.

Table 3.3: Results of first and second survey, store averages

	Survey 1				Survey 2			
	Control		Treatment		Control		Treatment	
	mean	sd	mean	sd	mean	sd	mean	sd
Job satisfaction	5.17	1.08	5.35	1.00	5.51	0.94	5.63	1.08
Task allocation:								
Ability	4.09	1.04	4.05	1.23	4.42	1.17	4.11	1.33
Preference	3.24	1.06	3.35	1.07	3.71*	0.94	3.29	0.99
Seniority	2.08	0.81	2.27	0.85	2.36	0.91	2.42	1.09
Fairness	5.28	1.37	5.55	0.92	5.60	0.75	5.33	1.20
Favouritism	2.09*	0.85	2.43	1.04	2.30	0.93	2.38	0.96
Efficiency	5.14	1.23	5.17	1.00	5.57	0.76	5.24	1.12
Observations	34-39		47-50		27-28		33-38	

*, **, and *** indicate that the difference between treatment stores and control stores in a given survey is statistically significant at the $p < .1$ level, $p < .05$ level, and $p < .01$ level, respectively.

Table 3.1 gives the exact wording of the questionnaire items. The exact number of stores varies across questionnaire items due to item non-response.

The left-hand side of Table 3.3 reports descriptive statistics of the first employee survey, conducted before the first tournament period. On average, employees indicate that fairness considerations and employee ability are the most important drivers of task assignment in their store. Employee preferences matter to a smaller extent while favouritism and seniority are not perceived as important drivers of task assignment. We find limited differences across first-period treatment and control stores. Only favouritism is perceived as slightly more important by employees in treatment stores as compared to employees in control stores (p -value 0.098). The average level of job satisfaction is also similar across treatment and control.¹⁴

¹⁴Across the different determinants of task allocation as listed in Table 3.1, between 34%

Descriptive statistics

Table 3.4 presents correlations between the store averages on the main items in the first survey. There is a rather strong positive correlation between the perceived importance of employee ability and the perceived importance of employee preferences in task allocation as well as between the perceived importance of seniority and favouritism. Not surprisingly, the perceived importance of favouritism and seniority both correlate negatively with fairness considerations. Interestingly, the extent to which task allocation is geared towards sales performance ('efficiency') is most strongly related to fairness considerations, and relates positively (negatively) to the importance of employee ability and employee preferences (favouritism and seniority) in task allocation. This could reflect that people are more productive under a fair task allocation. Another interpretation is that an efficient allocation is considered to be fair. Job satisfaction is most strongly related to the perceived importance of fairness and is also positively correlated with the importance of employee ability and employee preferences.

The right-hand side of Table 3.3 gives the outcomes of the second survey, conducted after the first tournament period. Compared to the first survey, job satisfaction is somewhat higher but not specifically among the treatment stores. The difference in the perceived importance of favouritism in task allocation between first-period treatment and control stores is no longer present in the second survey. We do find that employee preferences are considered less important for task assignment among stores that did participate in a first-period tournament than among stores that did not participate.

A concern in evaluating the effects of the team incentive on job satisfaction and task allocation is that only part of the employees completed the surveys, which may lead to selection effects. When this selection is related to stores' assignment to the treatment or control group, it yields biased estimates of the

and 44% of variation in responses to the first survey across employees is explained by store-fixed effects. Furthermore, restricting attention to stores that were in the control group during the first experimental period, correlations of average responses at the store level between the first and second survey range between 0.27 and 0.60. Hence, our measures of task allocation probably contain some measurement error through individuals' subjective assessment, but do seem to capture the underlying actual task allocation.

Table 3.4: Correlations between drivers of task allocation in first survey, store-level averages

Task Allocation	Ability	Preference	Seniority	Fairness	Favouritism	Efficiency
Preference	0.560***	1				
Seniority	0.118	0.073	1			
Fairness	0.222**	0.310***	-0.406***	1		
Favouritism	0.008	0.020	0.746***	-0.401***	1	
Efficiency	0.372***	0.423***	-0.333***	0.773***	-0.283***	1
Job Satisfaction	0.290***	0.300***	0.029	0.352***	-0.090	0.318***

*, **, and *** indicate that the coefficient is statistically significantly different from zero at the $p < .1$ level, $p < .05$ level, and $p < .01$ level, respectively.

Descriptive statistics

treatment effects. For the first survey, self-selection is unlikely to be problematic. At that point, employees were not yet aware of the upcoming experiment, so that self-selection cannot be based on assignment to treatment or control. Panel A in Table A.1 in the Appendix shows that first-survey respondents are older than non-respondents, have a longer history with the company, and work more hours, suggesting that employees with a stronger connection to the firm were more likely to complete the survey. Separating this by first-period assignment, we find indeed that across the treatment and control groups a comparable set of employees completed the survey. A similar pattern arises in the second survey as shown in Panel B. Furthermore, survey attrition is not significantly related to first-period treatment assignment. Conditional on answering the first survey, 38% of the employees in the control group answered the second survey against 34% in the treatment group.

Aggregating the survey data to store level, Table A.2 shows that stores with and without respondents to the first survey are comparable in terms of observable characteristics. For the second survey, we do find some differences between stores with and without survey respondents. Stores that performed relatively well before the experiment, that are headed by an older and more experienced manager, and that have a larger team are more likely to have at least one employee responding to the survey. Comparing the differences between stores with and without survey across the treatment and the control group, we find that these patterns arise in both groups (not reported for brevity). Hence, both the individual-level and the store-level data show that self-selection into the first and second survey appears to be unrelated to treatment assignment.

Self-selection can also be related to stores' task allocation. Again, this type of self-selection is most problematic when it differs between stores in the treatment and the control group. As assignment to treatment and control was stratified by the response to the first survey, self-selection into the first survey is unlikely to affect our estimated treatment effects. Furthermore, we can analyse whether self-selection into the second survey conditional on first-survey responses differs

between treatment and control stores. Panel A in Table A.3 reports the average response to the key questions in the first survey, comparing employees who only responded to the first survey with employees who responded to both surveys, separated by first-period assignment. This shows no substantial differences between employees who only responded to the first survey with employees who responded to both surveys. None of these differences differs significantly between the treatment group and the control group. Panel B in Table A.3 reports average responses to the second survey comparing employees who answered only to the second survey with employees who answered both surveys. In the control group, employees who completed both surveys indicate significantly higher importance of seniority and favouritism in their stores' task allocation compared to employees who only respond to the second survey. A similar, but less pronounced pattern arises in the treatment group. Comparing these differences across the treatment and control group, only for ability we find a statistically significant difference (p -value 0.06). Employees in the control group who answered only to the second survey indicate lower importance of ability than employees who answered both surveys, while the reverse holds in the treatment group. Assuming that the non-participation of respondents in the first survey is unrelated to treatment assignment, this implies that we may underestimate the treatment effect on the importance of ability in task allocation in the estimations at the employee level.

In Table A.4, we report similar figures aggregated at store level, comparing stores where at least one employee answered to each of the surveys with stores where none of the employees responded to one of the surveys. At the store level, we find that none of the differences between these types of stores differs significantly between the treatment and the control group. All together, the available evidence suggests that self-selection into the surveys does not affect our estimates.

3.5 Results

The first column of Table 3.5 gives the results of estimating (3.2). The estimated treatment effect of participating in a tournament is a reduction in performance by 0.6 percentage points. This effect is precisely estimated, with a standard error of 0.8 percentage points. Hence, the 95 percent confidence interval of the average treatment effect lies between -2.1 and 1.0 percentage point. In Column 2 of Table 3.5, we separate the average treatment effect by tournament period. In both periods, the estimated effect is small and statistically insignificant, and the difference between the estimated treatment effects for the two periods is small as well. In the third column, we estimate the effect of participating in the second tournament period separately for the 15 stores who had also participated in the first period, to establish whether there are carry-over effects of participating in the first tournament to performance during the second tournament period. We find no statistically significant carry-over effect, suggesting that having participated in the first tournament period does not affect stores' response during the second tournament period.¹⁵

The absence of a positive treatment effect is not due to low power. Given our difference-in-difference design, the power of our analysis depends on serial correlation across observations within stores (Bertrand et al. 2004). Using the period before the experiment took place (weeks 1 – 36), we regress performance on store- and week-fixed effects (as we also do in the main analysis). The residuals of this regression have a standard deviation of 0.12, as well as an estimated first-order auto-correlation of 0.19 and effectively no higher-order auto-correlation. We are not aware of an exact way to determine the power of an analysis in the presence of serial correlation. It is possible to give upper and lower bounds, though. In the absence of correlation across observations within stores, our design would allow us to detect an effect of 1.9 percentage point with a power of 0.8. On

¹⁵Within a tournament period, there is limited variation in the estimated treatment effect across weeks. In the first (second) half of the first tournament period, the estimated treatment effect is -1.8% (1.5%). In the first (second) half of the second tournament period, the estimated treatment effect is -0.2% (-1.3%). None of these coefficients differs significantly from zero.

Table 3.5: Estimation performance effect (intermediate weeks discarded)

	Dependent variable: Performance			
Treatment	-0.006 (0.008)			
Treatment in period 1	-0.004 (0.013)	-0.004 (0.013)		
Treatment in period 2	-0.008 (0.014)	-0.004 (0.016)		
Carry-over effects	-0.013 (0.022)			
Placebo-treatment				0.002 (0.011)
Store-fixed effects	yes	yes	yes	yes
Week-fixed effects	yes	yes	yes	yes
Observations	5274	5274	5274	3880
Stores	108	108	108	108
within R^2	0.565	0.565	0.565	0.615

Standard errors clustered at the store level in parentheses.

* and ** indicate that the coefficient is statistically significantly different from zero at the $p < .1$ level and $p < .05$ level, respectively.

the other hand, if all observations within stores would have a 0.19 correlation, we could detect an effect of 3.3 percentage points with 0.8 power. As only a small subset of observations within stores are correlated, the power of our analysis will be closer to the former than to the latter.¹⁶

Despite the random assignment, it could be that relatively many of the stores that participated in the first tournament period experienced a positive shock to performance just before that period. During the tournament period, performance may have moved back to normal levels, giving rise to a downward bias in the estimated treatment effect. To assess this possibility, we pretend a tournament took place in the period before the first real tournament, i.e. in weeks 31 to 36 in

¹⁶As shown by Bertrand et al. (2004), clustering standard errors at store level corrects for any serial correlation within stores. Consistent with the presence of weak serial correlation in the data, clustering increases the standard errors of the estimated treatment effects to a limited extent. An alternative way to handle serial correlation is to remove the time dimension from the data. Hence, we collapse our data into three periods, by taking average store performance before the experiment, during the first experimental period, and during the second experimental period. Running the difference-in-difference estimation using only these three periods, we obtain very similar results: the estimated effect of the treatment equals -0.004, with a standard error of 0.008.

Results

our data-set (see Figure 3.4).¹⁷ Thereto, we construct a dummy that takes value 1 in these six weeks for all stores that take part in the first tournament period. Column 4 of Table 3.5 gives the results of estimating the effect of this ‘placebo-treatment’, dropping all weeks afterwards. The estimated effect is very close to zero. Hence, the stores that participated in the first tournament period did not experience a positive shock to performance in the weeks before the tournament.

The absence of a positive average treatment effect on performance does not necessarily imply that the team incentive did not affect behaviour of managers and employees. The team incentive may have induced them to try out new ways of improving performance, for instance by making changes to task assignment in stores. Such attempts could be successful in some stores but fail in others. If so, we would observe an average treatment effect close to zero accompanied by a relatively large standard error. Comparing the standard errors on the actual treatment in Column 2 with the standard error on the placebo-treatment in Column 4 (which are all based on six-week periods), we find that the standard errors on the actual treatment are only slightly higher. Hence, the zero average treatment effect does not mask a large increase in the heterogeneity of store performance.¹⁸

Next, we use the survey data to assess whether employees perceived a change in task allocation in the first tournament period in response to the treatment. Table 3.6 gives the estimated effects of participating in a tournament in the first period on task allocation within teams, estimated at the store level. In contrast to our predictions, we do not find that our treatment increased the importance of employee ability in allocating tasks. The point estimate is negative, but not significantly different from zero. For the other considerations, we also find negative point estimates, all of them insignificant except for the importance of employee preferences in task allocation. Estimating these effects at the individual

¹⁷We leave out the week before the first tournament period (week 37), as that may pick up an effect of the announcement of the tournament.

¹⁸All results discussed carry over when we focus only on the effects in the first week or the first three weeks of each treatment period.

employee level yields a similar picture, as presented in Table 3.7. Summarising, while the reduced emphasis on fairness, favouritism, and in particular employee preferences in allocating tasks is in line with our predictions, this should have been accompanied by an increased emphasis on employee ability. Instead we find an insignificant negative effect. Overall, our findings suggest that the introduction of team incentives had little effect on the allocation of tasks within the teams.¹⁹

Table 3.6: Task allocation estimates, store level

Task allocation:	Ability	Preferences	Seniority	Fairness	Favouritism
Treatment	-0.243 (0.324)	-0.562* (0.287)	-0.136 (0.267)	-0.450 (0.321)	-0.300 (0.292)
Store-fixed effects	yes	yes	yes	yes	yes
Period-fixed effects	yes	yes	yes	yes	yes
Stores	60	60	60	60	60
within R^2	.009	0.101	0.015	.039	.028

Standard errors clustered at the store level in parentheses.

Dependent variables measured on a 7-point Likert scale, see Table 3.1 for the exact wording of the survey questions.

* and ** indicate that the coefficient is statistically significantly different from zero at the $p < .1$ level and $p < .05$ level, respectively.

Table 3.7: Task allocation estimates, worker level

Task allocation:	Ability	Preferences	Seniority	Fairness	Favouritism
Treatment	-0.352 (0.286)	-0.750** (0.325)	0.114 (0.223)	-0.412 (0.325)	-0.006 (0.386)
Worker-fixed effects	yes	yes	yes	yes	yes
Period-fixed effects	yes	yes	yes	yes	yes
Workers	91	91	92	90	92
within R^2	0.013	0.067	0.091	.027	.009

Standard errors clustered at the store level in parentheses.

Dependent variables measured on a 7-point Likert scale, see Table 3.1 for the exact wording of the survey questions.

* and ** indicate that the coefficient is statistically significantly different from zero at the $p < .1$ level and $p < .05$ level, respectively.

Table 3.8 reports the estimated effects of the treatment on employees' job satisfaction at the store level and at the worker level. The first and third column

¹⁹In line with this interpretation, we also find a small and insignificant treatment effect on employees' perception regarding the 'efficiency' of their store's task allocation (point estimate of -0.108 ; standard error of 0.239).

Discussion

show that the average treatment effect is small and statistically insignificant. This average treatment effect may mask heterogeneity across treated stores, in particular between stores that won their tournament and stores that did not win. However, we cannot estimate the effect of winning a tournament by simply including a dummy for stores that won their tournament, as performing relatively well may also affect job satisfaction in the absence of an tournament incentive. Hence, in order to differentiate between the effects of winning the tournament and the effect of attaining relatively high performance, we determine ‘winners’ and ‘losers’ of a pseudo-competition among stores that were part of the control group in the first tournament period. The pseudo-competition was conducted as follows. We assigned the control stores to groups of three in the same manner as we did with the treatment stores for the actual tournament. Next, we determined for each group of control stores the ‘winning’ store, based on stores’ cumulative performance during the first tournament period. The second and fourth column of Table 3.8 give the treatment effect for winning and non-winning stores separately. The first coefficient gives the treatment effect on non-winning stores, which is statistically insignificant both in the store-level estimation and in the worker-level estimation. Hence, participating in a tournament without winning it does not affect job satisfaction significantly compared to stores that did not participate and performed relatively poor as well. The second coefficient shows that job satisfaction goes down in control stores that outperformed two similar control stores during the tournament period, and significantly so in the store-level estimation. One explanation is that the higher performance is due to higher employee effort. Winning an actual tournament mitigates this effect, as seen by the third coefficient, although the effects are not statistically significant.

3.6 Discussion

Overall, our results show that the introduction of the team incentive neither affected team performance nor task assignment within teams. These results are

Table 3.8: Job satisfaction

	Dependent variable: Job satisfaction			
	Store level		Worker level	
Treatment	0.212 (0.333)	0.154 (0.366)	0.328 (0.307)	-0.356 (0.374)
Best in group		-1.241** (0.608)		-0.599 (0.588)
Best in treatment		0.575 (0.707)		0.281 (0.648)
Store/employee-fixed effects	yes	yes	yes	yes
Period-fixed effects	yes	yes	yes	yes
Stores / workers	53	53	59	59
within R^2	0.055	0.189	0.030	0.066

Standard errors clustered at the store level in parentheses.

Dependent variable measured on a 7-point Likert scale, see Table 3.1 for the exact wording of the survey question.

* and ** indicate that the coefficient is statistically significantly different from zero at the $p < .1$ level and $p < .05$ level, respectively.

in contrast to the results of earlier studies in the literature. An important question is: why? One way to address this question is to compare studies along important dimensions, and we will do so in the remainder of this section. However, before we delve into this, we would like to stress that the number of existing empirical studies providing causal estimates on the effects of team incentives is still quite small. In building up a body of evidence, it should be no surprise to sometimes find contrasting evidence (Antonakis 2017).

Our result on the effect of team incentives on task assignment is in contrast to Bandiera et al. (2007) and Burgess et al. (2010), who find that supervisors directed more competent workers towards the incentivized tasks. In contrast to these earlier studies, we do not observe the tasks actually performed by workers. Instead, we use subjectively reported drivers of store’s task assignment. This indirect method may underestimate changes in actual task allocation. Alternatively, the bonus offered may have been too low to induce changes in task assignment, which is in line with the absence of an overall treatment effect. For managers, the monetary incentive offered was comparable to the incentives offered in Burgess

Discussion

et al. (2010), but considerably lower than the incentive offered in Bandiera et al. (2007), as discussed in detail in subsection 3.2.2.

In our earlier field experiments with comparable designs and rewards conducted in other retail chains, we found average treatment effects on performance varying from 0%, 1.5%, to 5%, the latter two statistically significantly different from zero (Delfgaauw et al. 2013, 2014, 2015). Friebe et al. (2017) report a significant increase in sales of about 3% after the introduction of a team bonus for meeting sales targets in a retail setting. Hence, in earlier work, similar incentive schemes did induce higher performance.²⁰ Furthermore, part of our treatment entailed the provision of relative performance feedback. Studies in various settings find that the provision of relative performance feedback alone, absent relative performance pay, induces higher performance (Azmat and Iriberry 2010, 2016, Blanes-i-Vidal and Nossol 2011, Bradler et al. 2016, Delfgaauw et al. 2013, Kosfeld and Neckermann 2011, Kuhnen and Tymula 2012). Barankay (2012) and Bandiera et al. (2013), in contrast, find negative effects of relative performance feedback on performance at work. This suggests that the bonus level alone cannot explain the absence of a positive average treatment effect.²¹

²⁰While the bonus in our field experiment is equal to about 2.5% of earnings over a 6-week period and paid to a third of the participating employees (i.e. those of the winning store in a group of three stores), in Delfgaauw et al. (2013) employees could earn a bonus of 3.8% of earnings over a 6-week period, but this bonus was paid only to a fifth of the participating employees as stores competed in groups of five stores. However, the bonus scheme in that field experiment also included a bonus of 1.9% for the employees of the runner-up. Note, however, that Delfgaauw et al. (2013) also studied a treatment where stores compete in groups of five without any monetary prizes, yielding effects on performance of similar size. In Delfgaauw et al. (2014), employees could earn a bonus of 5% of monthly pay when the store would outperform a benchmark by a small margin and a bonus of 10% when outperforming the benchmark by a large margin. It turned out to be hard to qualify for the bonus: less than 11% of the stores earned a bonus, with half of these winning the low bonus. In Delfgaauw et al. (2015), teams entered elimination tournaments lasting a maximum of two times four weeks, with expected earnings of about 2% of monthly earnings, with prize money ranging from 1.2% to 6% of monthly earnings. Lastly, in Friebe et al. (2017), about 40% of the employees received a bonus of on average 4% at least once, implying an increase in total wage compensation of about 2%. We conclude that, by and large, the strength of the monetary incentives in these closely related studies is comparable to those in our study. Moreover, there is no clear relation between the strength of the incentive and the performance effect, which ranges from a null finding in Delfgaauw et al. (2014), to an increase in sales growth of about 5 to 7 percentage points in Delfgaauw et al. (2013), a 1.5% increase in the average number of products per customer in Delfgaauw et al. (2015), to a 3% increase in sales in Friebe et al. (2017).

²¹Several studies show a non-monotonic relation between the level of incentives and performance, where performance is lower for weak incentives than in the absence of incentives (Gneezy and Rustichini 2000, Gneezy and Rey-Biel 2014). Hence, it could be that the average treat-

Another possible explanation for the absence of an effect of our treatment is that the competitive element in our incentive design did not strike a chord due to the all-female composition of the teams. Gneezy et al. (2003) find in a lab experiment that females respond less to competitive incentives than males. Subsequent studies show that this gender difference in the response to competition depends on the specific task and the environment (Niederle and Vesterlund 2011). Studying a competitive business game played in groups of three students, Apesteguia et al. (2012) finds that all-female teams perform worse than any other team in terms of gender composition. However, Delfgaauw et al. (2013) implemented similar tournaments among stores of another retail chain and found that the treatment effect increases significantly in the fraction of female employees, provided that the store manager is female (which holds for all stores in the current study). Hence, this suggests that the all-female team composition likely does not drive the lack of response.²²

According to the company's management, the employees were actively engaged in the tournament. In some stores, the weekly rankings were eagerly awaited. However, the company's management perceived that many employees faced difficulties translating their engagement into higher sales, possibly due to a lack of skills to recognize and act on sales opportunities. Arguably, if employees are uncertain about how to increase performance, incentives may have little effect, at least in the short-run. After the current experiment, the company decided to invest in commercial training of its employees. Furthermore, it adopted an incentive scheme based on individual performance, which (unfortunately for us) was implemented in all stores at once.

ment effect would have been higher if we had only provided relative performance information. Still, our earlier experiments suggest that the current level of bonus pay can induce positive treatment effects.

²²Ceiling effects could be another explanation. If many stores perform as good as it gets, there is no room for improvement. Figure 3.1 suggests otherwise, however. As a further check we ran regression (2) adding the interaction between the first-period treatment dummy and the stores' average response to the question on efficiency in the first survey. This yields a negative coefficient for the interaction, suggesting that stores with more room for improvement respond relatively stronger to the team incentive. However, the effect is not statistically significant.

3.A Appendix

Table A.1: Characteristics of first and second survey (non-)respondents, by treatment

A: Survey 1						
	Overall		Control		Treatment	
	Non-Resp.	Resp	Non-Resp.	Resp	Non-Resp.	Resp
Tenure	7.28*** (7.71)	9.34 (8.48)	7.31** (7.82)	9.60 (8.78)	7.24** (7.64)	9.12 (8.24)
FTE	0.44*** (0.33)	0.54 (0.30)	0.47*** (0.33)	0.58 (0.28)	0.41*** (0.33)	0.51 (0.32)
Age	35.56*** (13.78)	39.49 (13.03)	35.00*** (13.34)	39.87 (13.02)	36.07** (14.17)	39.17 (13.09)
Observations	473-483	254-257	224-227	117-118	249-256	137-139
B: Survey 2						
	Overall		Control		Treatment	
	Non-Resp.	Resp	Non-Resp.	Resp	Non-Resp.	Resp
Tenure	7.83 (8.23)	8.71 (7.20)	7.96 (8.46)	8.68 (7.12)	7.72 (8.03)	8.73 (7.32)
FTE	0.46** (0.33)	0.54 (0.30)	0.48*** (0.32)	0.60 (0.29)	0.44 (0.33)	0.48 (0.31)
Age	36.40** (13.80)	39.23 (12.76)	36.25 (13.62)	38.49 (12.68)	36.52* (13.97)	39.87 (13.13)
Observations	589-601	138-139	277-281	64	312-320	74-75

Mean of each variable with standard deviation in parentheses.

The number of observations varies due to partial missing data.

*, **, and *** indicate that the difference in means between non-respondents and respondents within a group is statistically significant at the $p < .1$ level, $p < .05$ level, and $p < .01$ level, respectively.

None of the differences between non-respondents and respondents differs significantly across the treatment and control group at the 0.1 level.

Table A.2: Store characteristics and survey response

	Survey 1			Survey 2		
	No response	A response		No response	A response	
	mean	sd	mean	sd	mean	sd
Prior performance	0.96	0.14	0.95	0.06	0.93*	0.08
Team size	7.11	3.38	8.08	3.28	6.88***	3.58
Aides	1.74	1.69	1.94	1.53	1.57*	1.77
Age (manager)	41.23	14.06	44.34	10.33	40.80**	10.44
Tenure (manager)	14.07	12.51	15.71	14.21	11.05***	15.55
FTE (manager)	0.90	0.08	0.88	0.09	0.88	0.09
Age (staff)	39.55	9.43	38.16	6.48	38.31	7.33
Tenure (staff)	8.99	5.62	8.64	4.81	8.04	4.60
FTE (staff)	0.48	0.15	0.48	0.12	0.47	0.12
Observations	19		89		42	66

Comparing stores with and without responses on a given survey, *, **, and *** indicate that the difference in means is statistically significantly different from zero at the $p < .1$ level, $p < .05$ level, and $p < .01$ level, respectively.

Table A.3: Average worker responses by survey participation and treatment

A: Survey 1					B: Survey 2				
		Control		Treatment		Control		Treatment	
Participants in survey:	1	1 & 2	1	1 & 2	2	1 & 2	2	1 & 2	1 & 2
Job satisfaction	5.52 (1.21)	5.11 (1.63)	5.33 (1.43)	5.38 (1.46)	5.47 (1.36)	5.53 (1.07)	5.95 (1.00)	5.40 (1.33)	
Task allocation:									
Ability	4.01 (1.67)	4.42 (1.53)	4.01 (1.46)	4.30 (1.83)	3.89* (1.78)	4.59 (1.26)	4.46 (1.56)	4.13 (1.44)	
Preference	3.15 (1.39)	3.27 (1.34)	3.16* (1.35)	3.65 (1.74)	3.22* (1.59)	3.76 (1.21)	3.23 (1.39)	3.43 (1.23)	
Seniority	2.11 (1.30)	2.36 (1.38)	2.35 (1.34)	2.13 (1.21)	1.72** (0.89)	2.67 (1.43)	2.12 (0.99)	2.55 (1.33)	
Fairness	5.52 (1.50)	5.36 (1.19)	5.47 (1.34)	5.53 (1.54)	5.89 (1.02)	5.40 (1.09)	5.38 (1.36)	5.21 (1.56)	
Favouritism	2.15 (1.34)	2.51 (1.65)	2.54 (1.56)	2.36 (1.50)	1.67** (0.84)	2.64 (1.58)	2.23 (1.18)	2.49 (1.41)	
Observations	56-73	37-45	81-93	40-47	15-18	30-45	20-26	35-47	

Means of each variable with standard deviation in parentheses.
Comparing employees with and without responses on a given survey, *, **, and *** indicate that the difference in means is statistically significantly different from zero at the $p < .1$ level, $p < .05$ level, and $p < .01$ level, respectively.

Table A.4: Store average responses by survey participation and treatment

A: Survey 1					B: Survey 2			
Respondents in survey:	Control		Treatment		Control		Treatment	
	1	1 & 2	1	1 & 2	2	1 & 2	2	1 & 2
Job sat.	5.17 (0.97)	5.18 (1.13)	5.49 (1.13)	5.28 (0.95)	7.00 (.)	5.46 (0.91)	5.00 (1.73)	5.69 (1.02)
Task allocation:								
Ability	3.81 (1.32)	4.21 (0.90)	3.97 (1.22)	4.09 (1.25)	6.00 (.)	4.36 (1.15)	4.83 (1.62)	4.00 (1.27)
Preference	3.30 (1.39)	3.21 (0.91)	3.31 (1.07)	3.38 (1.09)	2.00 (.)	3.77 (0.90)	2.73 (0.83)	3.37 (0.99)
Seniority	1.83 (0.55)	2.19 (0.89)	1.97* (0.74)	2.43 (0.88)	2.00 (.)	2.38 (0.92)	2.03 (0.30)	2.48 (1.15)
Fairness	5.46 (1.95)	5.21 (1.06)	5.91** (0.67)	5.36 (0.98)	6.00 (.)	5.58 (0.76)	5.57 (0.43)	5.29 (1.28)
Favouritism	1.66** (0.45)	2.27 (0.92)	1.94** (0.95)	2.68 (1.01)	2.00 (.)	2.31 (0.95)	2.13 (0.51)	2.42 (1.01)
Observations	9-12	25-27	15-17	32-33	1	26-27	3-5	30-33

Means of each variable with standard deviation in parentheses.

Comparing stores with and without responses on a given survey, *, **, and *** indicate that the difference in means is statistically significantly different from zero at the $p < .1$ level, $p < .05$ level, and $p < .01$ level, respectively.

Chapter 4

Goal Setting and Raising the Bar

Joint with Max van Lent

4.1 Introduction

People often set goals. For example dieters commonly set a target weight, runners aim for a certain time, and managers set goals for employees in the form of targets. Using a series of field experiments, the psychologists Locke and Latham (1979) were the first to provide evidence that goals help to increase performance.¹ More recently, goal setting has also been studied by management scientists and economists. Economic theory papers have shown how (non-binding, non-monetary relevant) goals can be used as reference points in order to increase performance for loss averse agents or hyperbolic discounters (see e.g. Suvorov and Van de Ven 2008, Hsiaw 2013, Koch and Nafziger 2011 and 2016, and Koch et al. 2014), and that meeting goals can lead to a sense of self-achievement that

¹Locke and Latham found that goals set by an outsider (a peer or a manager), goals set in cooperation, and self-set goals can all lead to a better performance as compared to not setting goals.

makes pursuing goals worthwhile (Gomez-Minambres 2012). A growing empirical literature tests the effects of goal setting on performance in the laboratory and in the field.

This paper examines whether goal setting can help to increase student performance in an academic course. Furthermore, we are interested to learn whether challenging students to be more ambitious by increasing the goal's difficulty can increase performance further. This is relevant given the widely held belief that many students should be more ambitious, and the increased use of study counselors and mentors whose job it is to motivate and advise their students.

We test the effects of motivating students to set goals and attempts to raise the goal's difficulty by means of a field experiment among 1092 first-year economics students. Each of these first-year students regularly has meetings with a mentor (who is a senior student). Mentors help students to get used to studying at a university, teach them study skills, help them with their (study) motivation, monitor their performance, and give suggestions in order to increase their study performance. We ran the experiment during the second of three individual meetings between students and their mentor. In one treatment (*goal treatment*) we instructed mentors to ask their students whether they had a specific grade goal in mind for the main course they participated in at that moment, and if not, whether they wanted to set a grade goal. In another treatment (*raise treatment*) mentors received identical instructions as in the goal treatment, and were in addition instructed to encourage students to raise their goal if deemed appropriate. We subsequently measured performance using the grades the student obtained for the course.

We find that students whose mentor was instructed to motivate students to set a goal perform 0.18 better on a 10-point scale (which is 10.3% of a standard deviation) than students in the control group. This effect is driven by students in the goal treatment dropping out less often than students in the control group. Students whose mentor was instructed to also ask students to raise their goal do not perform significantly different from the control group.

Introduction

Setting goals can have adverse effects such as a narrow focus and even unethical behavior (see Ordonez et al. 2009). One novelty of this paper is that we can accurately measure some of these adverse effects. A concern in our setting is that students may increase effort and performance on the course for which they set a goal by simultaneously substituting away effort from the other course they take at the same time. We estimate the effect of the treatment on performance in the other course, and do not find such a negative effect. This implies that motivating students to set a goal is actually performance increasing overall.

In an attempt to uncover the mechanisms that drive the treatment effects we study heterogeneous treatment effects. We test whether there are heterogeneous effects of the treatments dependent on the student's prior study results, the mentor's experience, mentor's gender, and a match between mentor's and student's gender. We find that motivating students to set goals increases performance mainly for students who were initially performing poorly. We do not find a significant difference in treatment effect for students with more experienced mentors, or for students that have the same gender as their mentor.

There is a rich literature in psychology studying goal setting and its effects on performance (see Locke 1996, Locke and Latham 2002, and Locke and Latham 2006 for literature reviews). Research in psychology groups goals in roughly three categories: goals set by an outsider, cooperatively set goals, and self-set goals. Our *goal treatment* and *raise treatment* come closest to self-set goals and cooperatively set goals, respectively. Further, the literature shows that other factors such as goal commitment, goal specificity, and how challenging the goal is are important predictors for the success of goals (see for example Hollenbeck et al. 1989, Locke 1996, and Seijts et al. 2004). Goals that are specific, measurable and difficult but attainable tend to improve performance most. Our finding that the attempt to raise goals decreases students' performance as compared to goal setting by the student may be explained by a change in commitment to the goal, leading to a decrease in (study) motivation and hence performance.

Our paper is related to a rapidly increasing number of experiments in eco-

nomics that study the effects of different types of goal setting on performance in various contexts. Experiments range from self-set goals to goals set by others. In some papers goals are combined with monetary incentives (see e.g. Goerg and Kube 2012, Dalton et al. 2015, Corgnet et al. 2015, 2018, and Brookins et al. 2017) and in other papers goals are set without monetary incentives (see e.g. Goerg and Kube 2012, Sackett et al. 2014, Brookins et al. 2017, and Clark et al. 2017). These studies typically find that when ambitious but attainable goals are set, goals increase performance, and more so when they are combined with monetary incentives. Our main contributions to this literature are that we incorporate goal setting in a one-on-one mentoring program, and that we investigate the effects of raising goals by increasing its difficulty in a cooperative manner.

Also closely related to this paper is the literature on (non-monetary) incentives for students in education. This literature considers a number of ways besides setting goals, in which students' performance can be increased.² Lavecchia et al. (2016) review studies of interventions in education designed to improve students' performance. The interventions target a wide range of behaviors, varying from too little focus on the future, overreliance on routines, student self-confidence and the information on and number of choices in education. Further, Sanders and Chonaire (2015) show that in education usually (very) small effect sizes are found. The effect we find from goal setting is around the median effect size found in the sample of Sanders and Chonaire.³

Goal setting by students has received a lot of attention from psychologists, see e.g. Ames and Archer (1988) and Schunk (1990). Many of these papers in the psychology literature have tested whether goal setting can increase students' performance (see also Linnenbrink 2005, Morisano et al. 2010, Bettinger and Baker 2014, Schippers et al. 2015, and Travers et al. 2015). These studies are concerned with goal related activities such as coaching, self-reflection and

²For example changes in the class size (see Angrist and Lavy 1999 and Bandiera et al. 2010), providing feedback to students (see Bandiera et al. 2015), and several financial and non-financial incentives (see Levitt et al. 2016).

³While the median effect size in Sanders and Chonaire (2015) is 10% of a standard deviation, our (almost) costless intervention has an effect of 9.3% of a standard deviation.

essaying. In our experiment we explicitly ask the student to set a grade goal.

This paper is organized as follows. In the next section we explain the experimental context and describe the data. In section 4.3 we present a simple theoretical framework and derive our hypotheses. In section 4.4 we explain the empirical strategy. Section 4.5 presents the descriptive statistics, section 4.6 the results followed by the conclusion in the final section.

4.2 Experimental Set-up and Data Description

4.2.1 Experimental Context

The experiment involved 1092 first-year students enrolled in several undergraduate programmes at Erasmus School of Economics in Rotterdam, The Netherlands during the 2014-2015 academic year. The year is divided into five blocks of eight weeks. In each block students take 12 study credits (ECTS) worth of courses. All courses that students take at this point are obligatory, hence all students within a study programme take the same courses. Our experimental treatments take place during the second block when students have their second individual meeting with their mentor.

Each first-year student has a mentor. Mentors are senior students and are randomly assigned to students enrolled in the same programme at the start of the academic year. All mentors are employed by the university and are paid a flat wage. Our study involves all 84 mentors, and each mentor has 10 to 15 students. Mentors regularly meet with their students, both in groups and individually. Both group and individual meetings are obligatory. The mentor-student meetings are intended to teach students study skills, monitor their motivation, and more generally to provide a point of contact within the university. Motivation and individual prospects are the primary subjects of the three individual mentor-student meetings held over the course of the academic year. All individual mentor-student meetings take approximately 30 minutes. The first individ-

ual mentor-student meeting takes place around the start of the academic year in September, while the second and third take place in November and January, after the results of respectively the first and the second block of courses have been released. Our treatments are administered during the second individual mentor-student meeting.

While the first meeting at the start of the academic year primarily serves to discuss the student's motivation and to detect possible issues, the second and third meetings serve to evaluate results and prospects of the students. Due to university rules and national legislation at the time of the experiment, students with a weak performance record may be better off dropping out before February, which is in the third block of courses. Dropping out on time results in minimal grant loss and additionally allows students to re-enroll in the same programme the following academic year, which students that otherwise fail to meet first year requirements are not allowed to do. Thus, the second meeting is a natural moment to look forward towards the rest of the academic year and to discuss what results are necessary in order to make it sensible for the student to continue their current study programme. The last individual meeting after the release of the results for the second block serves mostly to determine whether it is better for the student to drop out given her motivation and study results.

Students participate in two courses in the second block, an introductory course in microeconomics worth 8 ECTS and a programme specific 4 ECTS course.⁴ Our treatment is focused on the microeconomics course. The course is taught in Dutch (824 enrolled students) and English (268 enrolled students). The Dutch and English versions are identical in all respects except for the lecturers and language spoken. The course follows a standard setup of three non-compulsory plenary lectures each week complemented by two compulsory tutorials taught by teaching assistants. The tutorials serve to review the course material, practise and discuss exercises, and in general to provide students an accessible way to

⁴Students enrolled in the Economics and Business Economics, Fiscal Economics, and Law and Economics programmes take an ICT course, while Econometrics students take a Calculus course.

obtain further explanation and clarification of the material. Tutorials are taught in 42 tutorial groups. One tutorial group consists of the students of two mentor groups. Examination of the course follows a standard format with two midterms counting 15% each and a written exam for the remaining 70%. For both midterms and the final exam students receive a grade on a 10 point scale, ranging from 1 to 10 with 10 being the best grade. In addition students could obtain a bonus, which was equal to at most half a point of the final grade, by participating in weekly online tests.

4.2.2 Experimental Design

Our experiment revolves around the second individual mentor-student meeting. We instructed a random subset of 54 of the 84 mentors to motivate their students to set a course specific grade goal during this mentor-student meeting. As discussed before, this second meeting is an excellent opportunity for such a discussion as its purpose is to reflect on past performance and consider what results for the current courses are necessary. This means that discussion of the progress of the current courses is natural, and a focus on microeconomics is expected since it is the most important course in the second block due to its weight in ECTS. Our treatment builds on this discussion.

During meetings with all mentors in the period between 22 and 31 October 2014, we informed the mentors that some of them would be expected to take a somewhat different approach to the second individual meeting. Selected mentors were sent instructions by e-mail about how to complement the discussion regarding the current courses one and a half week before the meetings. The instructions were accompanied by a simple flow diagram (see Appendices 1 and 2).⁵ All 54 selected mentors confirmed that they understood the instructions.

Randomly selected mentors were instructed to ask students whether they have

⁵Note that all mentors (including those assigned to the control group) received an extensive training and instruction from the mentor coordinator regarding how to conduct the mentor-student meetings. This reduces the chance that our instruction to the treatment mentors either signalled the importance of the meetings, or offers a guideline about how to structure the meeting.

a specific goal in terms of a grade in mind for microeconomics, and if so to elicit that grade goal. If the student did not have a grade goal in mind, the student was asked whether she wanted to set a goal on the spot, again eliciting the goal set. Students were free not to set a goal. Mentors were instructed to write down their evaluation of the goal of the student after the meeting, evaluating the student's goal as either "too easy", "doable" or "too hard".⁶ The description of the treatment so far describes the *goal treatment*. Thus in this treatment, mentors are asked to induce their students to set themselves a specific grade goal for the main course in the second block.

A second group of mentors were randomly selected to perform the *raise treatment*. In the raise treatment mentors implement the goal treatment but are in addition requested to attempt to raise the goal (if any) set by the student when deemed appropriate. If the mentor described the goal as "doable" or "too easy" the mentor was instructed to challenge the student by asking whether the student shouldn't be more ambitious and aim for a higher grade, specifically the student's self-set goal + 1 (e.g. if the student's goal was to get a 6 the mentor suggested aiming for a 7). The raise treatment serves to determine whether raising self-set goals can (further) improve study performance. Figure 4.1 illustrates the similarities and differences between the goal and raise treatment using a flowchart.

We chose to elicit a grade goal instead of other course related goals for multiple reasons. First, the final grade is (one of) the most important motivations to study for many students, hence students might find it more useful to set grade goals as compared to other goals. Second, choosing an output goal (the final grade) instead of an input goal (e.g. study hours) leads to lower measurement error because we cannot perfectly measure study hours. Finally, a grade goal is specific and measurable, which are important factors that influence the success of a goal (Locke and Latham 2002).

⁶The information that the mentor can use in order to evaluate the difficulty of the goal are the students' grades prior to the experiment, and the experience that the mentor has with the student based on one earlier individual meeting and several group meetings.

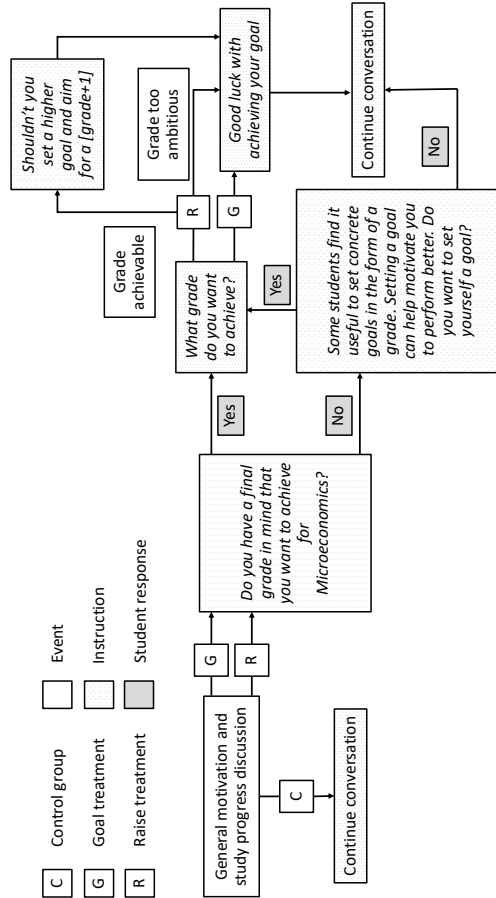


Figure 4.1: Flowchart of treatments

Mentors were asked to record the outcome of the meetings on a form. Mentors record whether students set a goal, what the goal is, and their estimate of the difficulty of the goal. In the raise treatment, mentors further record whether they asked students to raise their goal, the size of the raise, and whether or not the student accepted this higher goal. The mentor's estimate of the difficulty of the initial goal allows us to compare students in the raise treatment whose goal was challenged with similar students in the goal treatment whose goal was not challenged but would have been challenged if they were in the raise treatment.

Besides the forms filled in by the mentors selected to implement the treat-

ments we obtain information on all the students from administrative data from the microeconomics course and the central administrative office. This gives us information on the student's performance in other courses, attendance of microeconomics tutorial sessions, gender, age, study programme, and mentor.⁷ From the administration office we further obtained the mentor's gender and whether the mentor had experience in mentoring in previous years.

Only the mentors and lecturers were aware an experiment was being implemented, although mentors were not explicitly told so. Our introduction to all mentors in a general mentor instruction meeting necessitated that we informed all mentors that some of them would be asked to implement a small change in the upcoming individual mentor-student meetings. However, those not sent specific instructions were not aware of the exact change implemented. We specifically instructed the mentors who were selected for a treatment not to talk to anyone regarding our request. Selected mentors may deduce the purpose of the research but were not informed beyond their own instructions provided in Appendices 1 and 2.

4.2.3 Assignment Procedure

The assignment of students to both treatments and the control group is randomized at the mentor level. Assignment at the mentor level was chosen in order to increase compliance and prevent contamination. With assignment at the student level, a given mentor would be charged with treating her students differently, in a random order over the talks (students select a timeslot), likely leading to mistakes. In addition to accidental non-compliance, student level assignment might also result in more selective non-compliance by mentors selecting the treatment for their student(s) that they think is most appropriate.

The assignment of mentors to treatment was randomized in a stratified manner as follows. First, given that the tutorial group has a large impact on student

⁷For students in Dutch study programmes who attended a Dutch high school, we also have highschool grades.

performance as it is the main instruction method for many students, we ensure that a tutorial group is always of mixed composition in terms of treatments and control. This serves to create similar conditions for students in all treatments, but comes with the risk of contamination because students from treatment and control are in the same tutorial group. Randomization takes place within the various study programmes offered by the school as the effect of treatment can differ by programme due to the selection of students in a programme and the difficulty of the other course offered. Finally, several teaching assistants teach multiple tutorial groups. We therefore enforce that classes taught by the same teaching assistant have an (even) mix of control and treatment groups.⁸ In doing so we ensure that teaching assistants who teach classes in two different study programmes have a mix of control and treatment groups. Our randomization procedure is illustrated in Figure 4.2.

4.3 Predictions

In our setting, motivating students to set a grade goal can affect performance for two reasons: it may serve as a reference point and discussing the goal with the mentor may lead to reputational concerns with respect to the mentor.

Setting a goal in the presence of their mentor can lead students to change their behavior because of reputational concerns towards the mentor. Not only does the student know whether he or she failed to meet the goal, by sharing the goal the mentor may also become aware. To be precise, motivating students to share their goal with their mentor may lead the student to be more persistent in trying to reach the goal, leading to higher effort and performance.

The grade goal is also a clear reference point that can provide a demarcation between success and failure (besides that of passing the course). As such goals can be used as a motivation device by present-biased students who are prone to postpone effort. Setting a goal that requires efforts now allows students to

⁸For example, if a teaching assistant teaches two groups he teaches four mentor groups of which at least one group is assigned to each treatment and at least one group is a control group.

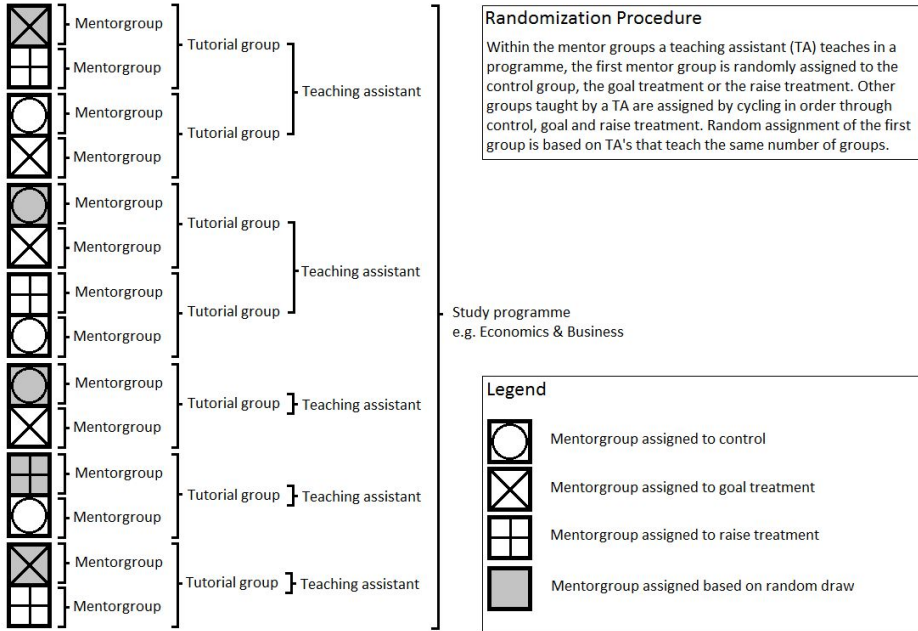


Figure 4.2: Randomization procedure

commit to effort now rather than postpone effort. Thus goalsetting can discipline the student and increase efforts. Note that non present-biased students will have no need for such a commitment such that the fact that a student sets a goal is revealing of their preferences.

The disciplining effect of the goal as a reference point is expected to result in higher effort and hence better performance. However, this need not be so if the goal set is too high. In that case students may at some point realize their goal is unattainable and subsequently lower their effort. The goal can be too high because of overconfidence (many students are known to be overconfident, see e.g. Bordalo et al. 2016), or especially in our experiment because students wish to impress their mentor. In addition in our experiment the goal can also be too high due to the raise treatment.

While the effects of goal setting are well studied in the literature, the effects

of proposing students to raise their goals are unstudied to our knowledge. We will now intuitively elaborate how motivating students to raise their goal can, but need not, increase study performance.

First note that reference points can serve as a disciplining device for present-biased students. However there is no guarantee that students set the optimal goal, and present biased students might also be tempted not to set too ambitious a goal. If so the goal can be raised without it becoming too difficult, leading proposing to raise the goal to improve study performance.

Second, the fact that students can choose whether to accept the proposed higher goal should attenuate a potential negative effect. Suppose the higher goal is too hard such that it would lower study performance (as explained above). Students will reject the proposal unless there is a (psychological) cost of doing so.⁹ Even if some students accept a proposed goal that is too ambitious not all will. The ability to choose attenuates a potential negative effect.

Third, the mentor's proposal to have the student increase her goal contains information. The student may believe that the mentor has superior knowledge of either the difficulty of the course, the ability of the student, or both. This may impact the student's effort decision, although it is unclear in which direction since this depends on the relation between effort and ability in the production function of which we do not want to make any assumptions.

Fourth, the effects of the proposal to raise may depend on how the raise is interpreted by the student. There is mixed evidence in the psychological literature regarding the origin of the goal on its effects, see e.g. Latham and Marshall (1982) and Hollenbeck et al. (1989). There is a potential pathway for self-set goals to have more of an effect than goals that are assigned or that are set in cooperation due to a higher locus of control. Motivating students to set a goal can be interpreted as a self-set goal. The proposal to raise the goal can be interpreted as a goal that is set in cooperation (or even as a goal that is assigned). Hence

⁹An example of a psychological cost is the cost of coming across as not ambitious towards the mentor.

the proposal to raise the goal may lead the student to value the goal differently.

Summarizing, while motivating students to set goals is expected to increase study performance, the effects of proposing students to raise their goals is less clear. If the proposed raise is small (relative to the current goal) then the proposal can increase performance.

4.4 Empirical Strategy

We estimate the effects of motivating students to set goals and attempts to raise students' self-set goals by estimating an intention-to-treat effect. The intention-to-treat measures the effect of a mentor being instructed to motivate students to set a grade goal, and in the raise treatment to attempt to raise students' goal. As becomes evident in the next section not all students are asked to set goals, for instance due to more pressing concerns in the meeting such as personal circumstances of the student. Also, not all students who are motivated to set a goal actually set a goal. We estimate:

$$P_i = \beta_0 + \beta_1 G_i + \beta_2 R_i + \beta_3 X_i + \varepsilon_i$$

where P_i is student i 's study performance, X_i a vector of control variables and ε_i the error term. G_i and R_i are treatment dummies indicating whether a student's mentor was assigned to the goal or raise treatment respectively. To be more precise on student performance, P_i is not the final grade of a student. The final grade for the course is composed of two midterm exams (both with weight 15%) and a final exam (with weight 70%). Since the mentor-student meeting is in the same week as the first midterm, we expect that students can hardly change their study behavior for the first midterm, and so we expect the treatment to only affect the later exams of the course. Hence we take as student performance a normalized combination of the second midterm and the final exam. Our performance measure

is hence calculated as $(0.15 \cdot \text{midterm2} + 0.7 \cdot \text{final}) / 0.85$.¹⁰ As a robustness check we also estimate the effect of our treatments on the final grade (i.e. including the first midterm).

There are two main reasons to include covariates in our regressions. First, since we assign treatment randomly conditional on the student's programme and teaching assistant we include dummies for the tutorial groups which subsume both these categories. Second, we include statistics on past study performance. Past study performance is highly predictive of present study performance and hence including measures of past performance reduces noise in the data, allowing for more precise estimates. We additionally include the student's gender, the mentor's gender and a dummy for the mentor's experience. In order to deal with inter cluster correlation, we cluster the errors of the regressions at the mentor level.

We assign students who do not complete the course a failing grade for those grade components that they do not complete. The highest failing grade a student could get is a 4.4 and the lowest a 1.0 at a 1 to 10 scale. In our estimations we focus on the lowest grade as the lowest failing grade is the grade that is actually given to students who do not pass the course. Further, for context consider that those who do not take the final exam but do take the second midterm score a 1.5 on average compared to the overall average of 5.7.

The effect of our treatments is composed of an effect on the intensive margin (performance of students who take the final exam) and an effect on the extensive margin (whether students participate in the final exam). We provide separate estimates for the effect on study performance for those students who complete the course, and for the effects on exam participation.¹¹

Ideally we would like to estimate the effect of actually setting a goal on student

¹⁰The weights assigned to the second midterm and the final exam in our performance measure, 0.15 and 0.70 respectively, are the same as the weights used in the composition of the final grade for students.

¹¹Note that all students (also those who do not take part in the final exam in December) still have the opportunity to participate in a resit exam which is administered in the summer. As a robustness check we later also estimate the effect of our treatments on the highest grade of the student including the possible resit exam grade.

performance using a local average treatment effect (LATE). However there are several reasons why a LATE estimation is infeasible in our setting. First, in order to estimate the effect on students who would be induced to set a goal by our treatment LATE estimation requires knowledge of the actual treatment a student receives, e.g. whether a student set a goal or not. We however don't know whether students in the control group actually set themselves a goal without their mentor explicitly motivating them.¹² Since our knowledge of actual treatment is perfectly correlated with treatment assignment we cannot distinguish so-called always-takers (the students who set a goal without the mentor motivating them to do so) from those influenced by our treatment. Thus a LATE estimation would amount to comparing the treated in the treatment group with the full control group. This by itself is expected to result in an underestimation of the effect, since if there is an effect it is also present amongst the always-takers in the control group. However, since mentors selectively motivate students to set goals (eg. the mentor may not discuss goals if the student has other issues to discuss) the selection into or out of treatment in the treatment group is biased towards including relatively good students into the treatment. This could lead to an overestimation of the treatment effect since a similar selection does not occur in the control group. As such we do not estimate local average treatment effects.

4.5 Descriptive Statistics

Our dataset contains information on 1092 students, 824 of whom are enrolled in a Dutch language programme with the remaining 268 students enrolled in an English language programme. Given that students are randomly assigned to mentor groups at the start of the academic year and the mentors are randomly assigned to treatment and control we do not expect to find any ex ante differences between the two treatment groups and the control group. Table 4.1 gives the

¹²We did not ask mentors in the control group whether their students set themselves goals. If we would have asked control group mentors whether their students set goals during the meeting, mentors could infer the purpose of our study. This could potentially lead mentors in the control group to alter their behavior.

descriptive statistics for the control (C), goal (G) and raise (R) group, as well as giving the p-value for two-sided comparisons of the means of these groups. Although the control and treatment groups appear to be comparable, there are some differences between the groups. Specifically, the characteristics of mentors of students in the treatment and control groups differ.¹³ Students in the control group are significantly more likely to have a female mentor whereas students in the raise treatment are more likely to have an experienced mentor.¹⁴ Furthermore treatment students in a Dutch language economics track (as opposed to students in an English language economics track) scored lower for the 8 credits accounting course in the first block than students in the control group, but there is no such difference regarding the mathematics course, which is more important for microeconomics.¹⁵ In the analysis we control for differences in observables.

Selection into or out of treatment is an issue affecting the generalizability of the results to the whole population. In our experiment there are three sources of selection out of the treatment. First, despite our best efforts to get all mentors to cooperate and ensure their understanding of the instructions, not all mentors assigned to treatment applied the treatment or took notes when administering the treatment. There are seven mentors for whom we do not have data about what happened during the individual student-mentor meetings. Anecdotal evidence suggests that some mentors have administered the treatment but did not record the results, while others did not administer the treatment. Thus this missing data forms a combination of measurement error and selection out of treatment. On average these mentors have less experience (5 out of 7 have no experience) than other mentors assigned to treatment (25 out of 47). Six of these seven mentors were assigned to the raise treatment, which was somewhat more demanding for mentors as they were asked to challenge their students goals. Mentors may also

¹³Note that there are only 84 mentors in the sample. In addition, at the time of the randomization the information on mentor characteristics was not available to us. Hence we could not stratify our randomization on mentor characteristics.

¹⁴We define an experienced mentor as a mentor who mentored students in earlier years.

¹⁵In Table 4.1 we tested for differences between control and treatment groups using t-tests. We obtain similar results if we use nonparametric tests.

Table 4.1: Descriptives by assigned treatment

	Control		Goal		Raise		C - G		C - R		G - R	
	mean	sd	mean	sd	mean	sd	p		p		p	
Age	18.72	1.71	18.63	1.39	18.75	1.27	0.57		0.85		0.39	
Female	0.29	0.46	0.30	0.46	0.30	0.46	0.87		0.95		0.93	
Female mentor	0.52	0.50	0.31	0.46	0.34	0.47	0.00***		0.00***		0.46	
Experience mentor	0.31	0.46	0.38	0.49	0.50	0.50	0.03**		0.00***		0.00***	
EC Accounting	6.45	1.61	6.13	1.51	5.96	1.73	0.06*		0.01**		0.31	
ECX Accounting	5.82	1.72	5.95	1.91	5.74	1.92	0.70		0.79		0.56	
EC Math	5.98	1.56	5.87	1.43	5.91	1.63	0.49		0.70		0.79	
ECX Math	6.48	1.51	6.64	1.63	6.66	1.38	0.58		0.51		0.96	
ET Matrix Alg.	6.48	1.94	6.13	1.88	6.02	2.38	0.32		0.29		0.79	
ETX Matrix Alg.	7.54	2.11	6.92	1.81	7.58	2.18	0.28		0.95		0.31	
ET Precalculus	6.46	1.50	6.04	1.78	5.72	2.20	0.17		0.05**		0.41	
ETX Precalculus	7.42	1.79	6.54	2.29	7.25	1.98	0.13		0.74		0.31	
ET Statistics	5.09	2.07	4.45	1.68	4.43	2.23	0.07*		0.13		0.96	
ETX Statistics	6.35	2.51	5.50	2.31	5.29	2.17	0.23		0.13		0.78	
Microeconomics	6.48	1.76	6.48	1.62	6.29	1.89	0.95		0.19		0.20	
Midterm I	4.89	2.25	4.80	2.10	4.78	2.23	0.58		0.55		0.94	
Attendance	10.66	2.98	10.66	3.15	10.71	2.85	0.98		0.81		0.84	
Dropout	0.14	0.35	0.11	0.31	0.14	0.34	0.21		0.94		0.26	
Max. Observations	389		367		336							

Students enroll in an economics (EC) track or an econometrics (ET) track in a Dutch or international (X) programme. Tracks in the Dutch and international programme are identical. Different tracks feature different courses, although some courses (e.g. Microeconomics) are common to all tracks.

Descriptive Statistics

feel more apprehension to administer treatment when their students had weak prior performance. However, compared to other students assigned to treatment, students of non-complying mentors do not differ in terms of prior performance. Since estimation of an intention-to-treat effect requires only knowledge of assigned treatment, the missing data does not have an effect on our estimates.

Second, there is some treatment dilution as mentors do not administer the treatment to all students. Mentors assigned to treatment ask students for their grade goal in 93% of the cases although they were instructed to administer the treatment to all. Moreover, mentors are selective in which students they target for treatment. Specifically, students who performed poorly in previous courses are less likely to be asked about their goals as is shown in Table 4.2. In cases in which mentors did not ask students about their goals they often noted a lack of time due to the necessity to discuss other issues. Also, conditional on receiving the data from the mentor, we find that more experienced mentors are less likely to administer treatment.¹⁶

Table 4.2: Students asked and not asked to set a goal within the treatment groups

	Not asked to set goal		Asked to set goal		p-value
	Mean	Observations	Mean	Observations	
Age	18.50	26	18.71	308	0.43
Female	0.30	37	0.27	483	0.69
Female mentor	0.22	37	0.34	492	0.14
Exp. mentor	0.62	37	0.46	492	0.06
EC Math	5.02	19	5.87	294	0.02
ECX Math	5.60	6	6.71	78	0.08
ET Precalculus	4.60	3	6.20	64	0.14
ETX Precalculus	5.50	2	7.43	22	0.24
EC Accounting	4.55	19	6.12	293	0.00
ECX Accounting	4.85	6	5.88	77	0.20
ET Matrix Alg.	4.47	3	6.35	63	0.10
ETX Matrix Alg.	6.70	2	7.85	22	0.46
ET Statistics	2.77	3	4.83	61	0.06
ETX Statistics	3.50	2	5.95	22	0.14

See Table 4.1 for an explanation of terms.

¹⁶Experienced mentors ask 90.5% of their students to set a goal, while non-experienced mentors ask 95% of their students to set a goal.

Table 4.3: Students asked and not asked to raise their goal in the raise treatment

	No raise proposed		Raise proposed		p-value
	Mean	Observations	Mean	Observations	
Age	18.79	84	18.59	75	0.29
Female	0.26	127	0.25	100	0.87
Female mentor	0.29	128	0.38	104	0.17
Exp. mentor	0.58	128	0.45	104	0.06
EC Math	5.63	70	6.04	74	0.13
ECX Math	6.32	21	7.24	17	0.03
ET Precalculus	6.26	7	6.97	3	0.73
ETX Precalculus	7.31	13	7.88	4	0.63
EC Accounting	5.69	69	6.24	75	0.06
ECX Accounting	5.25	21	6.31	17	0.10
ET Matrix Alg.	6.31	7	6.40	3	0.97
ETX Matrix Alg.	7.58	13	8.80	4	0.30
ET Statistics	5.36	7	7.40	3	0.22
ETX Statistics	5.44	13	6.47	4	0.37

See Table 4.1 for an explanation of terms.

In the raise treatment, mentors were instructed to attempt to raise the student's goal when they deemed the goal to be either too easy or doable. Of the 193 students setting a goal in the raise treatment 163 set a goal that met this requirement. However, mentors attempt to raise the goal in only 95 of these cases (58%), including all 47 cases where the goal is deemed too easy.¹⁷ Overall students who are asked to raise their goal have slightly higher grades than those not asked, although differences are largely insignificant.¹⁸ See Table 4.3 for more descriptives of the comparison between students asked and not asked to raise their goal.

It is of interest to note that 270 of the 492 students (55%) asked to set a goal already had a grade they wanted to achieve in mind. Comparing those students who had and had not yet a goal in mind, we see that those students who have a goal in mind are on average the better students (in terms of GPA), see also Table 4.4. The fact that many students already have a grade goal in mind implies that

¹⁷In addition, there are 9 instances where the mentor asks a student to raise the goal even though she estimated the goal to be too difficult.

¹⁸The low number of observations for the Dutch econometrics courses is due to the fact that three of the four Dutch econometrics mentors assigned to the raise treatment failed to provide data.

any effect of our treatment comes either from those who previously did not have a goal, or from the fact that students make the goal known to their mentor. The average initial goal set by the student is 6.9, a histogram of the goals set is shown in Figure 4.3. As expected higher (lower) goals are more likely to be deemed too hard (too easy) to achieve for the student by their mentor.

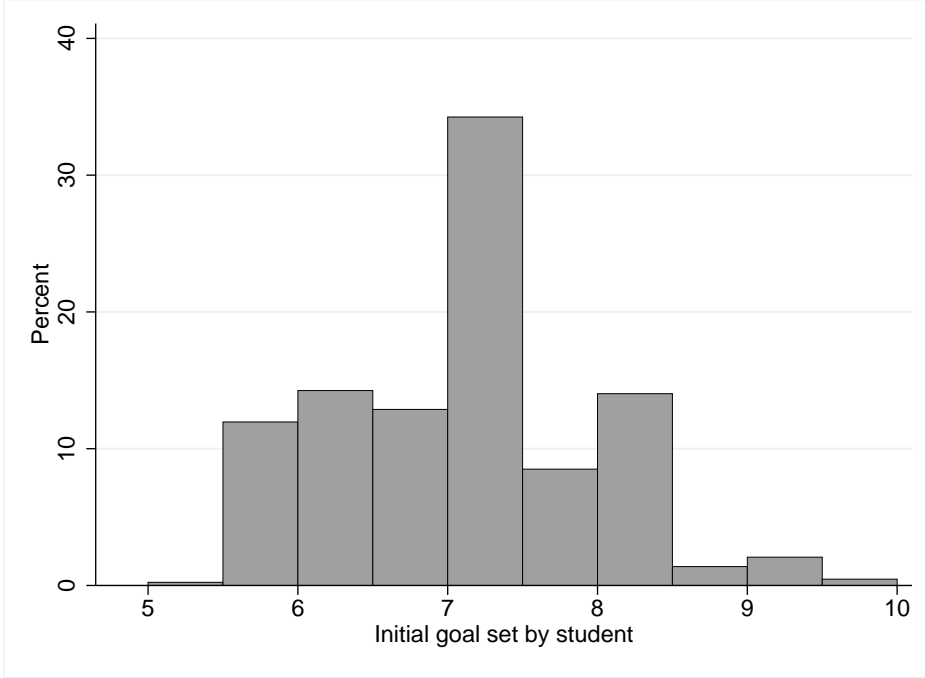


Figure 4.3: Initial goals set by students

Mentors appear to be able to gauge goal difficulty, as a regression of the difference between the final grade achieved and the initial goal set (both treatments) on the estimate of the mentor shows in Table 4.5. Goals that were expected to be too difficult were not achieved on average whereas goals that were too easy are indeed beaten by a significant margin. Furthermore all point estimates of the judgment categories differ significantly from each other, indicating that mentors differentiate well between the three categories. On average students failed to meet their goal by 0.4 of a point, which suggests that mentors are slightly overop-

Table 4.4: Students who have a goal in mind when asked

	Goal in mind		No goal in mind		p-value
	Mean	Observations	Mean	Observations	
Age	18.78	177	18.61	157	0.23
Female	0.27	265	0.27	255	0.90
Female mentor	0.30	271	0.36	258	0.16
Exp. mentor	0.53	271	0.41	258	0.01
GPA	6.06	256	5.71	238	0.02

See Table 4.1 for an explanation of terms.

Table 4.5: Mentor estimates are accurate

	grade - goal
Goal too hard	-1.507*** (0.261)
Goal too easy	0.337*** (0.412)
Baseline: doable	-0.446*** (0.261)
Observations	406
R^2	0.111

Standard errors in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

timistic about their students. There may be a concern that mentors assigned to the raise treatment are more likely to report that they expect the student's goal to be too difficult, in order to avoid challenging the students to raise their goal. We test whether the distribution of the mentor's estimates of the students' goal differs across treatments. We do not find evidence of such an effect.

Also 12% (59) of the 492 students asked for a goal do not set a goal. While students are more likely to set a goal if they have a female mentor there are no significant differences between those setting and not setting a goal in terms of past performance as shown in Table 4.6. Of all students asked to raise their goal half accept a higher goal. Again there is no significant difference in terms of past study results, but students are less likely to accept a raise from more experienced mentors (p-value 0.03), see Table 4.7. Furthermore, the level of the

Table 4.6: Students who set and do not set goals when asked to set a goal

	Set Goal		No goal set		p
	Mean	Observations	Mean	Observations	
Age	18.74	265	18.58	43	0.48
Female	0.27	424	0.27	59	0.94
Female Mentor	0.35	433	0.22	59	0.05
Exp. Mentor	0.47	433	0.41	59	0.39
EC Math	5.85	254	5.99	40	0.60
ECX Math	6.67	73	7.16	5	0.49
ET Precalculus	6.28	57	5.57	7	0.32
ETX Precalculus	7.62	20	5.55	2	0.19
EC Accounting	6.12	252	6.17	41	0.85
ECX Accounting	5.96	72	4.78	5	0.18
ET Matrix Alg.	6.28	57	7.00	6	0.39
ETX Matrix Alg.	7.89	20	7.40	2	0.74
ET Statistics	4.90	55	4.18	6	0.38
ETX Statistics	6.03	20	5.20	2	0.62
Midterm I	4.77	426	4.47	55	0.32

See Table 4.1 for an explanation of terms.

initial goal set has no influence on the acceptance of a suggested raise of the goal. Of the students who were asked to raise their goal we see that 50 percent (i.e. 52 students) reject the goal the mentor proposed. The average goal proposed does not differ (two-sided p-value of 0.62) between students who accept and reject the goal.

Finally, there are some differences between students participating in Dutch language and international (English) programmes. The programmes are identical in all respects, featuring the same courses, study materials, and examinations, but have a somewhat different application procedure. Students applying for a Dutch language programme need a high school diploma meeting the requirements, whereas those applying for an international programme are additionally selected based on a motivation letter. This suggests that students in the international language programmes might be better motivated. Consistent with this, Table 4.8 shows that, with the exception of an accounting course, students enrolled in international programmes scored significantly higher in the courses completed prior to the experiment. In our analysis we control for programme enrolment

Table 4.7: Students who accept and reject raise when asked to raise their goal

	Reject raise		Accept raise		p-value
	Mean	Observations	Mean	Observations	
Age	18.53	38	18.65	37	0.60
Female	0.22	50	0.28	50	0.49
Female mentor	0.46	52	0.29	52	0.07
Exp. mentor	0.56	52	0.35	52	0.03
EC Math	5.93	37	6.15	37	0.57
ECX Math	6.78	9	7.75	8	0.10
ET Precalculus	4.70	1	8.10	2	.
ETX Precalculus	8.85	2	6.90	2	0.34
EC Accounting	6.52	38	5.95	37	0.13
ECX Accounting	6.12	9	6.51	8	0.70
ET Matrix Alg.	4.60	1	7.30	2	.
ETX Matrix Alg.	9.90	2	7.70	2	0.33
ET Statistics	6.10	1	8.05	2	.
ETX Statistics	7.25	2	5.70	2	0.54
Initial goal	6.60	52	6.53	52	0.62

See Table 4.1 for an explanation of terms.

and additionally provide separate estimates of the effects within the Dutch and international programmes.

4.6 Results

We first provide the total effect of the treatments, imputing a failing grade for students who did not complete the course. We then provide the results for stu-

Table 4.8: Previous course results in Dutch and international programmes

	International		Dutch		p
	Mean	n	Mean	n	
EC Math	6.58	172	5.92	531	0.00
EC Accounting	5.84	170	6.18	529	0.02
ET Matrix Alg.	7.39	72	6.23	162	0.00
ET Statistics	5.84	71	4.67	158	0.00
ET Precalculus	7.14	71	6.10	164	0.00

See Table 4.1 for an explanation of terms.

dents who complete the course before turning our attention to the results for students who did not complete the course.

4.6.1 Total effect

We estimate the total effect of the treatments by imputing the highest and lowest possible grade that would result in failing to pass the course for those graded aspects of the course that the student did not complete. As discussed in section 4.4, we focus on the case in which we impute a missing grade as 1.0, as this appears to be the most relevant case.

Table 4.9 gives the result of the intention to treat estimations. We find a weakly significant ($p < 0.1$) positive effect of 0.18 of a grade point (i.e. 10.3% of a standard deviation) for students in the goal treatment and an insignificant negative effect of the raise treatment. The positive effect of assignment to the goal treatment is in line with our hypothesis that setting goals improves student performance. The insignificant negative effect in the raise treatment shows that attempts to raise a goal backfire, resulting in performance similar to students in the control group. The results are estimated separately for men and women in columns 3 and 4 of Table 4.9. We see that men hardly respond to the treatments, and the effects are (mainly) driven by women. Women in the goal treatment show a substantial positive effect of 0.39. There is also a significant negative effect of -0.27 of a grade point of the raise treatment. Columns 5 and 6 of Table 4.9 provide separate estimates for students in Dutch and international programmes. These estimates show a widely divergent response to the treatments in these two groups, with Dutch programme students responding positively to the goal treatment while students in international programmes respond to the raise treatment in a strongly negative manner. This difference in response may be explained by differences in the initial study motivation of students (students in the Dutch program are admitted based on only a high school diploma while International students are besides selected on grades also selected on the basis

of a motivation letter). Finally, in the second column of Table 4.9 we show the results excluding the students of mentors who did not hand in a form. These students are less likely to have been subjected to treatment, therefore the second column may give an indication of the effect size when compliance rates are higher than they were in our experiment. Recall from the descriptive statistics section that students of mentors not handing in forms do not differ from those of other mentors in the treatment groups in observable characteristics. The positive effect of the goal treatment is stronger in these estimates while the effect of the raise treatment is again negative but insignificant. The overall picture that emerges from the intention to treat estimates is that setting goals can help, but attempting to raise goals undoes any positive effect of setting goals and may even result in worse performance as compared to students that were assigned to the control group.

The total effect of our treatments discussed above consists of two effects. First the treatments may have an impact on the students who complete the course inducing them to alter their efforts. This is the effect on the intensive margin. Second, our treatments may affect the decision to participate in the course. These two effects cannot be interpreted separately as this risks the confusion of selection effects on the extensive margin for the effects of treatment on the intensive margin. We turn to these two effects now.

4.6.2 Intensive margin

Table 4.10 gives the intention-to-treat estimates for those students that complete the course (i.e. the students who participate in the final exam). The results are largely in line with the overall estimates provided in Table 4.9. We find that the overall positive effect of being assigned to the goal treatment is no longer significant for the students who complete the course. But results in the female and Dutch programme subsamples are similar and significant. The results confirm the overall impression that motivating students to set goals can improve student

Table 4.9: Intention to treat effect: total effect

Grade	Overall	Ex. 7 mentors	Male	Female	Dutch	International
Missing grade=1.0						
T: Goal	0.181** (0.091)	0.278*** (0.111)	0.152 (0.111)	0.394** (0.187)	0.339*** (0.115)	-0.011 (0.125)
T: Raise	-0.165 (0.105)	-0.123 (0.107)	-0.031 (0.127)	-0.265* (0.155)	0.019 (0.118)	-0.428** (0.177)
Observations	967	878	682	285	724	243
Mentorgroups	84	77	84	82	60	24
R ²	0.619	0.628	0.627	0.679	0.618	0.627

Standard errors in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 4.10: Intention to treat effect: intensive margin

Grade	Overall	Ex. 7 mentors	Male	Female	Dutch	International
Missing grade=1.0						
T: Goal	0.112 (0.077)	0.193** (0.077)	0.069 (0.092)	0.361** (0.178)	0.265*** (0.090)	0.011 (0.125)
T: Raise	-0.147 (0.092)	-0.117 (0.095)	-0.007 (0.103)	-0.262* (0.151)	0.054 (0.091)	-0.428** (0.177)
Observations	940	884	661	279	697	243
Mentorgroups	84	77	84	82	60	24
R ²	0.609	0.616	0.626	0.674	0.605	0.627

Standard errors in parentheses.
* $p < .1$, ** $p < .05$, *** $p < .01$

performance, and that proposing to raise the goal has an insignificant negative effect on performance.

4.6.3 Extensive margin

The results above indicate a positive effect of the goal treatment on course performance for those students who complete the course. Differences between the total effect estimates and the estimates on the intensive margin may be due to selection effects induced by the treatments. For instance, our treatments may affect course completion by creating greater commitment. To study the effects of our treatments on course completion we estimate a linear probability model in much the same way as above.

In the control group 6.2 percent (24) of the students having attended at least 3 sessions dropped out of the course. This dropout rate is lowered by 2 percentage points on average in the goal treatment as can be seen in Table 4.11 providing the intention to treat estimates on the dropout rate.¹⁹ The results on the dropout rate are similar to those on course performance given course completion in that the goal treatment again has a positive effect on performance (i.e. dropout decreased) whereas the raise treatment does not have an effect but has an oppositely signed coefficient. In contrast to the effect on the course grade however, the reduction in dropouts is concentrated among men rather than women. This is most likely due to the fact that women have a substantially lower baseline dropout rate than men (6.8% for men compared to 2.8% for women in the control group, two-sided p-value 0.12).

In our setting students cannot retake the entire course. However, at the end of the academic year there is a resit for the final exam. In order to see whether dropping out of the course more often is undesirable (and dropping out less often is desirable), we estimate the final grade taking the maximum of the final exam grade from the regular exam and the final exam grade from the resit. Using this

¹⁹No estimates on subsamples for the Dutch and international programmes are provided as there are too few dropouts from the international programmes, resulting in collinearity.

Table 4.11: Intention to treat effect: dropout

Dropout	Overall	Ex. 7 mentors	Male	Female
T: Goal	-0.021*** (0.008)	-0.029*** (0.008)	-0.216** (0.110)	- 0.016 (0.013)
T: Raise	0.004 (0.008)	0.000 (0.008)	-0.000 (0.012)	0.009 (0.011)
Observations	967	878	682	285
Mentorgroups	84	77	84	82
R^2	0.165	0.180	0.172	0.267

Standard errors in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 4.12: Intention to treat effect: maximum grade including resit

Grade	
T: Goal	0.063 (0.075)
T: Raise	-0.192** (0.091)
Observations	940
Mentorgroups	84

Standard errors in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

approach we find that the positive effect of the goal treatment becomes smaller, and the negative effect of the raise treatment becomes stronger and significant, see Table 4.12. This implies that it is unlikely that students in the raise treatment strategically dropout in order to receive a higher grade using the resit exam.

4.6.4 Further results

In this section we present a number of additional analyses that shed light on further questions. To start, we consider whether there is a heterogeneous effect of treatment due to differences in ability. We measure ability by taking the average of the grades achieved in the first block, and centering this grade average by subtracting the overall mean average score of 6.2 (std. dev. 1.65). We then interact the ability measure with students' treatment assignment. The intention-

to-treat estimates in Table 4.13 show that students who performed better in previous courses respond less to the goal treatment. Thus our intervention had a stronger positive effect on weaker students than it did on top students. There is no such heterogeneous effect regarding the raise treatment.²⁰

Second, as we have seen above, students set higher goals in front of a female mentor (two-sided p-value of 0.012). This suggests that the motivation to set a goal may differ depending on the gender of the mentor. It is thus natural to analyze whether our treatments have heterogeneous effects depending on the mentor's gender. Overall there is no sign of a heterogeneous treatment effect depending on the gender of the administering mentor as shown in Table 4.14.

The desire to impress a member of the opposite gender may result in a heterogeneous treatment effect based on the gender combination of mentor and student. Estimates on male and female students reveal no such significant interaction effect, although the signs of the point estimates do conflict in the expected manner which is consistent with such an effect. In the subsample of Dutch students there is a clear interaction effect of gender on treatment response. The point estimates suggest that having a female mentor completely reverses the effect of the treatments such that the raise treatment has an insignificant negative impact when administered by a male mentor but a positive impact when administered by a female mentor. Male and female mentors have the same effect on Dutch programme students in the goal treatment. Furthermore, no gender differences are found in the sample of international students, although this may be due to a lack of power.

Another possible channel through which treatment may be affected is the experience of the mentor. More experienced mentors may project more authority or may be better able to fit the treatment into the conversation in a natural manner. A heterogeneous treatment effect based on the mentor's experience may also speak to a possible channel through which setting a goal in the individual

²⁰For students who attended a Dutch high school we also have high school grades. If we use high school grades (as a measure of ability) for this subsample we find a qualitatively similar result but a decrease in power because of the smaller sample size.

Table 4.13: ITT interaction treatment with GPA in preceding block

	Overall	Ex. 7 mentors	Male	Female	Dutch	International
T: Goal	0.172** (0.075)	0.255*** (0.076)	0.123 (0.090)	0.389* (0.198)	0.323*** (0.094)	0.048 (0.109)
T: Goal \times GPA	-0.103** (0.050)	-0.108* (0.055)	-0.169*** (0.061)	-0.103 (0.128)	-0.142** (0.061)	-0.418 (0.087)
T: Raise	-0.085 (0.079)	-0.063 (0.084)	0.029 (0.096)	-0.109 (0.142)	0.093 (0.081)	-0.361** (0.142)
T: Raise \times GPA	-0.053 (0.058)	0.039 (0.067)	0.020 (0.076)	0.080 (0.099)	-0.016 (0.063)	0.208* (0.112)
GPA	1.064*** (0.263)	0.992** (0.434)	1.248*** (0.242)	0.607 (0.065)	0.556** (0.243)	1.035*** (0.308)
Observations	914	831	644	270	675	239
Mentorgroups	84	77	84	81	60	24
R^2	0.627	0.632	0.682	0.568	0.633	0.624

GPA is the mean centered GPA of students calculated over all courses they participated in prior to microeconomics.
Standard errors in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 4.14: ITT interaction treatment with gender of mentor

Grade	Overall	Ex. 7 mentors	Male	Female	Dutch	International
Female mentor	-0.0002 (0.134)	-0.010 (0.136)	-0.053 (0.148)	-0.146 (0.300)	0.159 (0.113)	-0.105 (0.282)
T: Goal	0.171 (0.113)	0.218* (0.117)	0.002 (0.128)	0.554** (0.262)	0.387*** (0.082)	0.007 (0.156)
T: Goal \times Fem. mentor	-0.198 (0.228)	-0.083 (0.212)	0.085 (0.241)	-0.711 (0.535)	-0.520** (0.258)	0.002 (0.427)
T: Raise	-0.178 (0.111)	-0.131 (0.121)	-0.121 (0.133)	-0.430** (0.214)	-0.091 (0.092)	-0.186 (0.271)
T: Raise \times Fem. mentor	0.070 (0.178)	0.033 (0.169)	0.262 (0.212)	0.336 (0.395)	0.449** (0.205)	-0.487 (0.292)
Observations	940	854	661	279	697	243
Mentorgroups	84	77	84	82	60	24
R^2	0.609	0.616	0.626	0.678	0.609	0.629

Standard errors in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Results

mentor-student meeting affects performance. If students care for their reputation in the eyes of their mentor they may value this reputation more if the mentor is more experienced. Taken combined, we expect more experienced mentors to strengthen the treatment effects. Recall however that we have already seen some evidence speaking against a better implementation of the treatments by more experienced mentors in the form of a weakly lower rate of inducement (two-sided p-value of 0.06) and goals that are approximately three tenths of a point lower.

The estimation results including a heterogeneous effect for experienced mentors is given in Table 4.15. We see that more experienced mentors do not have a significant positive impact on the raise treatment overall. Men in the goal treatment with a more experienced mentor show an increase in performance whereas there is no such increase in performance when they have an inexperienced mentor. Female students in the goal treatment seem to do worse with an experienced mentor compared to an inexperienced mentor, although this effect is not significant. Although there appears to be a significant negative interaction effect of mentor experience on assignment to the goal treatment for international students, the effect of having a more experienced mentor implementing the treatment is no different from that of an inexperienced mentor. These results show that generally the students do not appear to be more responsive to more experienced mentors. Although we suggested that students might be more responsive to a more experienced mentor due to reputational concerns, this finding does not discredit the reputation explanation for why goals can improve performance. For instance, students may not distinguish that much between different mentors, or value their reputation in the eyes of their mentor more based on other factors such as the mentor's own ability.

We are fully aware that there are two different effects of setting a goal in our goal treatment. The effect of thinking about a goal and setting a goal for oneself, and the effect of sharing this goal with the mentor. Thinking about a goal may create a reference point with which to compare one's performance as posited in the literature. This may result in higher performance by giving positive utility if

Table 4.15: ITT interaction treatment with experience of mentor

Grade	Overall	Ex. 7 mentors	Male	Female	Dutch	International
T: Goal	0.067 (0.100)	0.202* (0.102)	-0.142 (0.124)	0.416* (0.225)	0.124 (0.140)	0.123 (0.091)
Experienced Mentor	0.087 (0.137)	0.047 (0.125)	-0.172 (0.133)	0.544* (0.280)	0.052 (0.124)	0.760** (0.303)
T: Goal \times Exp. Mentor	0.129 (0.200)	-0.032 (0.192)	0.606*** (0.229)	-0.250 (0.424)	0.292 (0.231)	-1.217*** (0.348)
T: Raise	-0.040 (0.132)	-0.088 (0.137)	0.124 (0.144)	-0.075 (0.264)	0.141 (0.137)	-0.254 (0.268)
T: Raise \times Exp. Mentor	-0.198 (0.167)	-0.073 (0.159)	-0.091 (0.183)	-0.502 (0.433)	-0.136 (0.188)	-0.862 (0.557)
Observations	940	854	661	279	697	243
Mentorgroups	84	77	84	82	60	24
R^2	0.610	0.616	0.629	0.676	0.606	0.630

Standard errors in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Results

the goal is met or exceeded and negative utility if performance falls short. Sharing the goal with the mentor may result in reputational concerns towards the mentor, in the sense that the mentor may evaluate the student based on her performance relative to her goal. Empirically our best way to learn whether sharing the goal with the mentor leads to the increased performance, is by considering when reputational concerns are likely to be stronger and comparing students in those situations with similar students who face weaker reputational concerns. We posit two such situations, facing a more experienced mentor who is likely better in evaluating students performance, and facing a mentor of the opposite gender who may be more valuable to impress. For neither of these situations we find a heterogeneous impact on the treatment effect, suggesting that either goals primarily work through creating a reference point for oneself or our proxies capture not enough variation in the strength of reputational concerns.

Furthermore, asking students to raise their goal can have effects through two different channels. The challenge of the goal itself can have an effect, i.e. if students get utility from reaching a goal they may adjust their effort if the goal becomes more challenging. Another channel is that asking the student to raise her goal can give the student information about her ability. Depending on the substitutability or complementarity of ability and effort this can lead to a decrease or increase in effort and performance. In order to learn if students adjust their effort because the proposal to raise the goal gives them additional information, we study heterogeneity in how much information a proposal to raise is expected to contain. Therefore we compare the treatment effects for students with experienced mentors with the treatment effects for students with unexperienced mentors. The reason is that we expect that students who have experienced mentors expect their mentor to be more able to know the student's ability, and hence the information aspect is stronger. We find that there is no significant difference in treatment effects for students who have experienced and unexperienced mentors. This implies that either the information component is small, or that there is not much heterogeneity in the extent to which the message contains

information by experience of the mentor.

One potential explanation for the fact that the raise treatment has a negative effect on performance is that the proposed raises are too high. In order to explore this we can make use of the fact that some student-mentor meetings take place after the first midterm. After the midterm, students as well as mentors have more information about what goals would be realistic for the student. Hence we expect goals as well as raises to be more realistic after the midterm than before. We test for the effects of differences in timing of the meeting, and find that timing of the student-mentor meeting does not affect students' willingness to set a goal, and neither does timing affect the probability that the student is asked to raise the goal. Finally, the timing of the meeting does not seem to have an effect on student performance.²¹

We find a consistent pattern in our data that female students respond stronger to our treatments than male students. At first sight this may seem surprising. There is a very rich literature on heterogeneous gender effects to monetary incentives and non-monetary incentives when there is a competition element. Many papers find that males respond stronger to competitive incentives or information about their (relative) ranking, see for example Gneezy et al. (2003), Barankay (2011), and Niederle and Vesterlund (2011), while others find that there is no gender difference, see Dreber et al. (2011) and Delfgaauw et al. (2013). An important difference between these incentives and the incentives in our treatments is that goal setting in this experiment does not have any competitive element, which might drive the gender effect in the current literature. Apicella et al. (2017) for instance find that males are more willing to compete against others than females, but find no gender difference when people compete against themselves. However, a recent paper by Clark et al. (2017) who study goal setting among university students using surveys, find that goal setting is more effective for male students. Similarly in a laboratory experiment Smithers (2015) finds that goals improve

²¹However, we have to note that students are allowed to pick a timeslot for the student-mentor meeting. Hence we cannot make any causal claims on the effects of timing of the meeting.

Conclusion

male but not female performance. Hence, the way in which goals are elicited (during a personal meeting or using surveys) may also have an impact on how effective goal setting is.

In our mean specification the outcome variable is a weighted average of the second midterm and the final exam. However, some students in both the control group and treatment groups had their mentor-student meeting shortly before the first midterm. Therefore the intervention could, for some students, have influenced even their first midterm grade. This is especially the case if students cram their study effort very close to the exam. Hence, as a robustness check we estimate our main specifications using the final grade that the students achieved. We find similar results using this final grade as the outcome variable.

A potential concern of goal setting in multitasking environments is that goals lead subjects away from other (non-incentivized) tasks. Our students take one other course at the same time as microeconomics. Doing our previous analyses with the other course's grade as a dependent variable shows that there is no evidence of such a substitution effect (these results are available on request). Hence, the positive effect of goal setting in this study is a net increase in performance.

4.7 Conclusion

We conducted a field experiment in order to test the effects of encouraging students to set goals and encouraging students to increase the ambitiousness of their goals during mentor-student meetings in a university study programme. We designed two treatments. In the *goal treatment* we instructed mentors to encourage students to set a grade specific goal. In the *raise treatment* we gave mentors the same instruction and in addition instructed them to raise this goal if deemed appropriate.

We find that students in the goal treatment perform better than students in the control group. Students whose mentor was assigned to the goal treatment score 0.18 gradepoints (i.e. 10.3% of a standard deviation) higher than students in

the control group. Students in the raise treatment perform similarly to students in the control group, although there are some indications that their performance is even lower. This is true in terms of both the dropout rate and the grades achieved that are conditional on participating in the final exam of the course. The null effect of the raise treatment is in line with the goal becoming unacceptable due to the raise, indicating that the size of the raise was too high.

An alternative explanation for the result that students in the raise treatment perform worse than students in the goal treatment is the nature of the goal. While in the goal treatment students set themselves a goal, a proposal to raise this self-set goal can be seen as a goal of a different kind, namely a cooperatively set goal (or even an assigned goal). Changing the nature of the goal can change the commitment of the student to the goal (see Hollenbeck et al. 1989), which implies that the intrinsic motivation (i.e. the utility gain when reaching the goal) changes between the two treatments. As a consequence students perform worse in the raise treatment than in the goal treatment. Further, if some students in the control group set themselves a goal then this could even lead to a lower performance of students in the raise treatment as compared to the control group.²² Finally, we find that students who performed poorly prior to the experiment benefit most from being motivated to set goals in a one-on-one meeting with their mentor.

It is interesting to learn whether these results are generalizable to other settings, for example manager-worker settings. Our findings may have implications for the optimal design of appraisal interviews. In order to generalize our findings we need more (experimental) evidence in other settings.

Our paper is (relatively) silent on the mechanisms that drive goal setting. It is interesting to learn to what extent present bias preferences and loss aversion, as is posited in economic theory papers as important drivers that make goals work, are predictors of the success of goal setting. For example our result that goal setting works mostly for initially poor performing students may be explained by

²²Our finding that 55% of students that are asked about goals already have a goal in mind supports this idea.

Conclusion

poor performing students having stronger present bias preferences, are more loss averse, or by the fact that extra effort more easily increases performance when initial performance is low. Van Lent (2017) takes a first step in disentangling these effects by measuring students' motivation, effort, and time preferences and relating this to the performance of students who are motivated to set goals in a randomized field experiment.

4.A Mentor instructions

4.A.1 Goal treatment

Dear X,

Following our introduction during the tutor instruction session, we request you to adjust the progress meetings with your students. Your participation contributes to research regarding the possibilities to increase students' study success by improving the tutor meetings.

The instructions regarding the progress meetings that you conduct in the week of 17th to 21st November follow.

After you have discussed the general motivation and study progress of the student, you are expected to ask some additional questions while discussing the current courses the student follows. These questions relate to Microeconomics. The intention is to motivate the students to set a goal. Ask the questions in italics.

Do you have a final grade in mind that you want to achieve for Microeconomics?

If YES: *What grade do you want to achieve?*

(if the student answers: I want to pass the course, then try to specify this, for example: Are you aiming for 5,5 or a 6?)

If NO: *Some students find it useful to set concrete goals in the form of a grade. Setting a goal can help motivate you to perform better. Do you want to set yourself a goal?*

If YES: *What final grade do you want to achieve for Microeconomics?*

If NO: continue the conversation as usual.

If the student set a goal:

Good luck with achieving your goal.

It is important that you follow the instructions as much as possible. However,

Mentor instructions

do try to incorporate the questions into the conversation naturally. Attached, you find a flowchart summarizing the script. You can use this flowchart to refresh your memory prior to the meeting.

We request you to complete the attached form after the meeting with each student. You can print this form yourself, or pick up a copy at H8-23 or H8-24. It can be useful to make notes. Please, read the form carefully before the meetings.

After the meetings we would like to receive the completed forms. The completed forms can be handed in at H8-23 or H8-24 or can be emailed to vanlent@ese.eur.nl or souverijn@ese.eur.nl. We request that you hand in the form at Friday 28th November at the latest.

For research purposes we request that you do not discuss these instructions with others.

If you have any questions, do not hesitate to contact us. You can find us in H8-23 or H8-24, and you can reach us at vanlent@ese.eur.nl, phone: 010 408 1793 or souverijn@ese.eur.nl, phone: 010 408 9038.

Max van Lent

Michiel Souverijn

P.S. Could you please confirm to us by email that you have received this email and that you have read the instructions.

4.A.2 Raise treatment

Dear X,

Following our introduction during the tutor instruction session, we request you to adjust the progress meetings with your students. Your participation contributes to research regarding the possibilities to increase students' study success by improving the tutor meetings.

The instructions regarding the progress meetings that you conduct in the week of 17th to 21st November follow.

After you have discussed the general motivation and study progress of the student, you are expected to ask some additional questions while discussing the current courses the student follows. These questions relate to Microeconomics. The intention is to motivate the students to set a goal. Ask the questions in italics.

Do you have a final grade in mind that you want to achieve for Microeconomics?

If YES: *What grade do you want to achieve?*

(if the student answers: I want to pass the course, then try to specify this, for example: Are you aiming for 5,5 or a 6?)

If NO: *Some students find it useful to set concrete goals in the form of a grade. Setting a goal can help motivate you to perform better. Do you want to set yourself a goal?*

If YES: *What final grade do you want to achieve for Microeconomics?*

If NO: continue the conversation as usual.

If you (as a tutor) think the goal (grade) set is achievable:

Shouldn't you set a higher goal and aim for a [grade +1]? [So if the student chooses a 6 as a goal and you think this is achievable, propose to aim for a 7.]

If the student set a goal:

Good luck with achieving your goal.

Mentor instructions

It is important that you follow the instructions as much as possible. However, do try to incorporate the questions into the conversation naturally. Attached, you find a flowchart summarizing the script. You can use this flowchart to refresh your memory prior to the meeting.

We request you to complete the attached form after the meeting with each student. You can print this form yourself, or pick up a copy at H8-23 or H8-24. It can be useful to make notes. Please, read the form carefully before the meetings.

After the meetings we would like to receive the completed forms. The completed forms can be handed in at H8-23 or H8-24 or can be emailed to vanlent@ese.eur.nl or souverijn@ese.eur.nl. We request that you hand in the form at Friday 28th November at the latest.

For research purposes we request that you do not discuss these instructions with others.

If you have any questions, do not hesitate to contact us. You can find us in H8-23 or H8-24, and you can reach us at vanlent@ese.eur.nl, phone: 010 408 1793 or souverijn@ese.eur.nl, phone: 010 408 9038.

Max van Lent

Michiel Souverijn

P.S. Could you please confirm to us by email that you have received this email and that you have read the instructions.

Summary

This thesis studies the ways in which incentives can change behaviour. The incentives studied take on a variety of forms, including target-based pay, piece-rates, tournaments and goalsetting. The variety of incentives is testament to the notion that different types of incentives have different strengths and weaknesses, and hence are appropriate in different settings. Among the settings studied in this thesis are manager-employee relationships, teamwork and schooling. Incentives are studied both theoretically and experimentally. The results of the studies in this thesis are summarized in this chapter.

Summary

In chapter 1 I study theoretically what incentives contracts are optimal when performance measurement is imperfect in the sense that performance measurement is not as granular as the underlying process. In the principal-agent model used the principal and agent can react to intermediate information regarding the value of production to which the performance measure is insensitive. The lack of granularity of the performance measure is potentially troubling because it could interfere with the possibility to adjust efforts to newly released information. For instance, if the value of output turns out to be higher than expected the agent optimally works harder subsequent to the revelation of that information. The non-granular performance measure could hamper the ability to elicit such activity by forcing compensation based on total performance including that prior to

the revelation of information. This potentially induces the agent to engage in gaming activities and impose additional costs on the principal if she wishes to induce the agent to change his behaviour.

It is shown that the problems discussed above can be avoided. Properly designed incentive contracts can elicit the efficient level of effort regardless of whether the performance measure is based on output or input and whether target- or piece-rate-based pay is employed. When the performance measure is output-based the principal optimally offers the agent a performance-compensation menu allowing the agent to select which option to work towards during the game. Note that the agent does not choose a contract to accept but is instead offered a contract which enumerates multiple options. The agent responds to new information during the game by selecting an option and adjusting efforts accordingly. Thus utilizing an output-based performance measure allows (forces) the principal to rely upon the agent's processing of new information.

In contrast to the 'front-loading' of contracting that occurs when an output-based performance measure is used, when an input-based performance measure is used the contract that ultimately determines the agent's pay is finalized at a later stage. An initial contract is signed but this is renegotiated after additional information becomes available. The initial contract takes a very specific form: it is set up to maximize value for one of the extremes of the value of efforts distribution. Thus if the chosen extreme of the distribution is realized no renegotiation is required. The rationale for this is as follows. Although the initial contract is necessary in order to protect the agent from extortion by the principal (who can withhold pay for efforts already provided) it also serves as the starting point for negotiation after the revelation of information. With the initial contract in hand the agent can engage in gaming activities that improve his payoffs if the new information is favourable to his machinations, while sticking to the initial contract when the new information is unfavourable. This limits the principal's ability to renegotiate the contract and hence the possibility of adjusting efforts to new information per se. A contract that can in principle be adjusted up- and

Summary

downwards will in practice never be subject to adjustment in both directions due to the agent's ability to speculate upon adjustment in one direction. It is therefore optimal to write a contract that allows for speculation in one direction only by focusing on an extreme of the information distribution. This both harnesses the agent's gaming in the proper direction and allows the contract to always be renegotiated. As a result only an extreme contract can implement the efficient outcome in case of an input-based performance measure.

In chapter 2 Josse Delfgaauw and I study the role of managers in assessing worker performance in a multitasking setting. By tying compensation to managerial performance evaluation managers are given authority over workers. This allows managers to influence what work their subordinates do, which they can use for the benefit of the organization or themselves. We develop a principal-supervisor-agent model with multitasking. The agent's compensation depends on an imperfect performance measure and the supervisor's assessment. We show that a manager can be used in much the same way as an imperfect performance measure can, leading to similar results with similar bias and monitoring ability or reliability. Thus biased supervision forms an alternative to using an imperfect performance measure. We further show that unlike multiple objective performance measures it is not straightforward to combine biased supervision with objective performance measurement. A supervisor can easily adjust her assessment depending on the context, hence depending on how the objective performance measures are being used. As a result combining biased supervision and objective performance measurement can lead to more congruency but is not guaranteed to do so.

We use the results to shed light on the discussion of assignment of managerial positions to specialists or generalists. We argue that while specialists may have better monitoring ability, they may also have more biased preferences. This gives a trade-off between the strength of subjective performance incentives and the distortion induced in the agent's efforts if objective performance measures are unavailable. Verifiable performance measures decrease both the cost of supervisor

bias and the benefit of higher monitoring ability. The effect on the former is typically stronger such that better performance measures make specialist supervisors relatively more attractive.

In chapter 3 Josse Delfgaauw, Robert Dur and I study the effects of a team incentive on team performance and division of labour by way of a field experiment in a retail chain. In teamwork performance depends not just on the efforts of individual team members but also on the task assignment within the team. Teams can engage in specialization by having teammembers perform the tasks they are relatively good at. However, team performance is not necessarily the only driver of task assignment as considerations such as fairness, seniority and favouritism may also play a role. Team incentives may focus the team on performance, bringing about more efficient task assignments.

We conduct a field experiment among stores of a national retail chain, introducing short term tournaments between randomly selected subsets of three stores. The sales-based tournaments were held among three participating stores and lasted for six weeks. Each employee of winning stores received a monetary prize equal to approximately three percent of monthly earnings. Both the sales and the results of surveys held before and directly after the tournaments are compared between participating and non-participating stores. We find no effect of the team incentive on sales. Surveys held in all stores before and after the incentive did not indicate an effect on task assignment in the participating teams.

Finally in chapter 4 Max van Lent and I study how to motivate students to perform better in an undergraduate class. We perform a field experiment among close to 1100 first-year students, motivating a random subset to set a grade goal. The students in the study all follow the same course and are all assigned to a mentor with whom they have regular (one-to-one) meetings. Treatment, if any, occurs during a one-to-one meeting with the mentor. The first treatment consisted of the mentors asking students for a grade goal. Students in this treatment are expected to do better than other students due to the positive effect of goalsetting. We find this group did better than the control group. In the

Summary

second treatment students who set a grade goal were challenged to set a more ambitious goal if the mentor evaluated the self set goal as easy or moderately difficult. While we expected this to push goals into the challenging zone and hence to further improve performance we in fact find a reversal of the positive effect of setting goals (the first treatment), such that students in this treatment perform similar to students in the control group.

Further research

The overarching theme in this thesis is that aligning interests, that is incentivizing all involved parties, is possible but by no means easy. A given incentive mechanism can work in one situation but backfire in another as different parts of the incentive structure interact with each other and the environment in which they are being used. It remains challenging for incentive theorists to predict what the reactions to an incentive will be. Sometimes seemingly minor changes in the incentive structure can result in big effects, while at other times high hopes are dashed by a muted response. We still need further understanding of how different components of the incentive structure such as incentive pay, norms, goals, team cohesion and others work together. This will help narrow down what works best in a particular environment and help explain the variety in incentive structures within industries.

Related to the above we should consider that the theoretical work that is being done in this field provides internally consistent conclusions that are relevant only if its premises are valid, in our case the assumptions being made regarding the incentive structure. The big question therefore is what the real world looks like and how incentive theory fits. While there is an increasing number of field experiments to test incentive theory the total number of studies is not that large and captures a somewhat haphazard slice of the economy. Field experiments are an important tool to show us how incentives work in specific instances, but do not provide a complete enough picture. Experiments tend to test a single

intervention in a specific situation, providing a less than complete picture. In general there seems to be a dearth of so called stylized facts regarding what types of incentives are used by what type of firms in what type of industries undergoing what type of transformations. More information on the prevalence of actual incentive structures provides a test of current incentive theory, suggests which theories are relevant and can show lacunae in the theory.

Samenvatting

Dit proefschrift bestudeert de wijze waarop incentives (prikkels) gedrag kunnen veranderen. Verschillende typen incentives zijn bestudeerd, waaronder op target gebaseerde beloning, stuksbeloning, toernooien en het zetten van doelen. De verschillende typen incentives hebben andere sterktes en zwaktes en zijn daardoor geschikt in verschillende omstandigheden. De omstandigheden die in dit proefschrift aan de orde komen zijn manager-werknemer relaties, teamwerk en studie. De incentives worden door middel van theorie en experimenten bestudeerd. In dit hoofdstuk volgt een samenvatting van de studies.

Samenvatting

In hoofdstuk een bestudeer ik welke contracten er theoretisch optimaal zijn wanneer het meten van de prestaties imperfect is in de zin dat de metingen minder vaak plaatsvinden dan het onderliggende prestatieproces. In het gebruikte principaal-agent model kunnen de principaal en agent reageren op tussentijdse informatie met betrekking tot de waarde van de productie van het werk van de agent. De prestatiemaatstaf meet echter de prestatie van de agent als geheel, dus zonder onderscheid te maken tussen prestaties voor en na het vrijkomen van deze informatie. Dit zou tot een probleem kunnen leiden omdat de principaal en agent de vrijgekomen informatie niet goed kunnen gebruiken om de inzet van de agent aan te passen aan de nieuwe informatie. Als bijvoorbeeld de waarde van de prestaties van de agent hoger dan verwacht blijkt te zijn zou het optimaal zijn dat

de agent harder werkt nadat de informatie is vrijgekomen. Een prestatimaatstaf met beperkte granulariteit over de tijd kan dit proces belemmeren doordat zo'n maatstaf beloning op de totale prestatie afdwingt, inclusief de prestaties voor het vrijkomen van betere informatie. Dit geeft de agent een prikkel om spelletjes te spelen met zijn inzet die de principaal op extra kosten kunnen jagen als de inzet van de agent veranderd dient te worden.

Ik laat zien dat het hierboven geschetste probleem voorkomen kan worden door goed ontworpen contracten. Goed ontworpen contracten incentiveren de agent tot het leveren van de efficiënte inzet ongeacht of de prestatimaatstaf op inzet of uitkomsten is gebaseerd of dat er sprake is van een contract op basis van targets of stuksbeloning. Als de prestatimaatstaf op uitkomsten gebaseerd is bied de principaal de agent optimaliter een menu van contracten. De agent kiest gedurende het spel voor welk van de opties hij zich inzet. Merk op dat de agent hierbij op nieuwe informatie kan inspelen door een andere optie na te jagen en zijn inzet te veranderen. De principaal kan (moet) dus vertrouwen op het verwerken van de nieuwe informatie door de agent.

Waar bij een op uitkomsten gebaseerde prestatimaatstaf van tevoren een uitgebreid contract wordt opgesteld wordt bij het gebruik van een op inzet gebaseerde maatstaf het uiteindelijke contract pas later volledig bepaald. Er wordt een initieel contract opgesteld, maar dit wordt heronderhandeld nadat er meer informatie beschikbaar komt. Het initiële contract heeft een specifieke vorm in de zin dat het geoptimaliseerd is voor een van de extremen van de distributie die bepaalt hoe waardevol de inzet van de agent is. Als de gekozen extreme de werkelijke situatie blijkt is er dus geen heronderhandeling nodig. De reden voor deze contractuele vorm is als volgt. Een initieel contract is noodzakelijk om de agent te beschermen tegen uitbuiting door de principaal, immers bij het opstellen van een contract nadat de extra informatie vrijkomt kan zij weigeren te betalen voor de inzet van de agent tot dat punt. Het initiële contract dient echter ook als uitgangspositie voor de heronderhandeling. Omdat de agent door middel van het initiële contract verzekerd is van een bepaalde uitkomst heeft hij de ruimte om

Samenvatting

te spelen met zijn inzet. Hij kan er bijvoorbeeld op gokken dat de waarde van zijn inzet heel hoog blijkt te zijn door alvast hard te werken. Als zijn gok goed uitpakt plukt de agent hier de vruchten van, en zo niet kan hij vasthouden aan het initiële contract. Dit verhindert de principaal om de agent ten alle tijden zijn inzet te laten aanpassen aan de beschikbare informatie. Een contract dat in principe ruimte biedt om omhoog en omlaag aangepast te worden zal in de praktijk nooit beide kanten op aangepast kunnen worden doordat de agent beter af is te speculeren op een aanpassing in een enkele richting. Het is daarom optimaal om het initiële contract zo op te stellen dat het slechts in één richting aangepast hoeft te worden. Dit zorgt ervoor dat de prikkel voor de agent om te speculeren op de uitkomst in goede banen wordt geleid en behoudt de mogelijkheid om het contract te heronderhandelen als er meer informatie beschikbaar komt.

In hoofdstuk twee bestuderen Josse Delfgaauw en ik de rol van managers in het beoordelen van prestaties in een omgeving waarbij arbeiders meerdere taken uitvoeren. Door beloning te koppelen aan de beoordeling van een manager verkrijgt een manager autoriteit over de arbeider. Dit stelt de manager in staat te beïnvloeden wat de arbeider doet, een macht die ten goede van de organisatie of ten goede van zichzelf kan worden ingezet. We ontwikkelen een principaal-manager-agent model met meerdere taken. De beloning van de agent hangt af van een imperfecte prestatiemaatstaf en de beoordeling van de manager. De manager heeft net als de prestatiemaatstaf een mogelijk andere weging van de verschillende taken ten opzichte van de principaal. We tonen aan dat een manager op een gelijke wijze als een imperfecte prestatiemaatstaf kan worden ingezet, met gelijke resultaten voor soortgelijke afwijkingen in de weging van taken en soortgelijke betrouwbaarheid en beoordelingsvermogen. Prestatiebeoordeling door managers vormt hiermee een alternatief voor het gebruik van een imperfecte prestatiemaatstaf. Het is in tegenstelling tot het gebruik van meerdere imperfecte prestatiemaatstaven echter niet eenvoudig om een imperfecte prestatiemaatstaf te supplementeren met prestatiebeoordeling door een manager. Dit komt omdat een vooringenomen manager haar beoordeling kan veranderen afhankelijk van de

context, oftewel afhankelijk van hoe de beoordeling door de principaal gebruikt wordt. Zij kan zo altijd een ‘goede’ beoordeling geven indien gewenst. Waar het combineren van meerdere imperfecte prestatimaatstaven leidt tot een minder imperfecte meting van prestatie hoeft dat dus niet het geval te zijn als een imperfecte prestatimaatstaf gecombineerd wordt met prestatiebeoordeling door een manager.

We gebruiken dit resultaat om meer licht te werpen op de discussie of de positie van manager aan specialisten of generalisten moet worden toevertrouwd. We gaan er vanuit dat specialisten beter zijn in het feitelijk beoordelen maar dat ze ook een sterkere mening hebben met betrekking tot het relatieve belang van de verschillende taken. Daarmee ontstaat een afweging tussen de kracht om de agent te beïnvloeden en de distortie die een specialist teweegbrengt ten opzichte van een generalist als er geen prestatimaatstaf beschikbaar is. De beschikbaarheid van een objectieve prestatimaatstaf verlaagt zowel de kosten van een bevooroordeelde manager als de baten van betere beoordelingscapaciteit. Omdat het eerste effect typisch sterker is dan het laatste zijn specialisten relatief aantrekkelijker.

Het derde hoofdstuk omvat een veldexperiment in een winkelketen naar de effecten van een teambeloning op teamprestaties en taakverdeling dat samen met Josse Delfgaauw en Robert Dur is uitgevoerd. Bij het werken in een team is niet enkel de individuele inzet van teamleden belangrijk maar ook de verdeling van de taken binnen het team. Door teamleden die taken uit te laten voeren waar ze relatief goed in zijn kunnen de teamprestaties geoptimaliseerd worden. De teamprestatie is echter niet noodzakelijkerwijs de enige reden om taken op een bepaalde manier te verdelen. Andere factoren zoals een eerlijke verdeling, senioriteit en vriendjespolitiek kunnen ook een rol spelen in de taakverdeling. Door het invoeren van een teambeloning kan het team zich meer gaan richten op de teamprestaties, waarvoor een efficiënte taakverdeling nodig is.

Het veldexperiment is uitgevoerd onder winkels van een nationale winkelketen, waarbij er kortlopende toernooien zijn georganiseerd tussen willekeurig gekozen

Samenvatting

groepen van drie winkels. De toernooien met elk drie deelnemende winkels duurden zes weken en werden beslist op basis van de omzet. Werknemers van winnende winkels ontvingen elk een financiële bonus van ongeveer drie procent van hun maandelijkse salaris. Van deelnemende en niet deelnemende winkels worden de omzetcijfers en de resultaten van enquêtes die voor en na de toernooien zijn gehouden vergeleken. We vinden geen effect van de teambeloning op de omzet. Uit de enquêtes blijkt geen effect op de taakverdeling binnen de teams die meededen met de toernooien.

In het vierde en laatste hoofdstuk bestuderen Max van Lent en ik hoe studenten gemotiveerd kunnen worden om beter te presteren. In een veldexperiment onder bijna 1100 eerstejaars studenten is een subset van studenten aangespoord om een doel te stellen in de vorm van een cijfer. Deze studenten volgen allemaal hetzelfde vak en hebben allemaal een mentor toegewezen gekregen met wie ze regelmatig (één-op-één) gesprekken voeren. Het is in de mentorgesprekken dat de experimentele interventie plaatsvindt. Er zijn twee interventies. In de eerste interventie vraagt de mentor de student om een doel in de vorm van een cijfer te stellen. De verwachting is dat studenten in deze groep beter presteren dan studenten in de controlegroep doordat het stellen van doelen motiveert om goed te presteren. Dit is ook wat we terug zien in de data. In de tweede interventiegroep worden studenten ook gevraagd een doel in de vorm van een cijfer te stellen. Als de student een goed haalbaar doel stelt wordt de student nu echter uitgedaagd een ambitieuzer doel te stellen. Hoewel we verwachtten dat dit de prestaties zou verbeteren omdat uitdagender doelen mogelijk meer aanzetten tot betere prestaties vinden we geen effect op de prestaties ten opzichte van de controlegroep. Het aanzetten tot ambitieuzere doelen deed de positieve effecten van het stellen van een cijferdoel teniet.

Verder onderzoek

Het overkoepelende thema in dit proefschrift is dat het mogelijk maar niet eenvoudig is om alle neuzen een kant op te krijgen door middel van economische prikkels. Een gegeven mechanisme kan in de ene omgeving goed werken maar in een andere omgeving perverse effecten hebben doordat verschillende onderdelen van de omgeving anders op elkaar ingrijpen. De exacte reactie op een prikkel blijft daarmee lastig te voorspellen. Soms resulteren ogenschijnlijk kleine veranderingen in grote effecten, terwijl op andere momenten hoge verwachtingen beloond worden met magere resultaten. We hebben een nog beter begrip nodig van hoe verschillende elementen als prestatiebeloning, normen, doelen, team samenstelling en anderen samen tot een resultaat leiden. Een beter begrip van de interacties helpt om beter te bepalen wat in een specifieke omgeving het beste werkt en kan helpen verklaren waarom er zulke verschillende vormen van beloningsbeleid zijn binnen industrieën.

Gerelateerd aan het bovenstaande moeten we in het achterhoofd houden dat de sterke theoretische resultaten in het onderzoeksveld gestoeld zijn op aannames, in het specifiek aannames rondom de beloningsstructuur. De grote vraag is dan hoe deze aannames aansluiten bij de werkelijkheid. Hoewel er in toenemende mate veldexperimenten worden gedaan om de theorie te toetsen aan de werkelijkheid is het aantal studies op dit vlak nog beperkt, en omvatten deze studies slechts een klein gedeelte van de gehele economie. Daarnaast geven de veldexperimenten wel weer hoe een specifieke prikkel in een specifieke omgeving werkt maar zijn de uitkomsten moeilijker te generaliseren naar een bredere omgeving. We missen in zekere zin het grote plaatje, het beeld van welk type mechanisme gebruikt wordt in welke situatie, in welke industrie en in welke combinaties. Meer data met betrekking tot het echte gebruik van incentive mechanismen zou ons in staat stellen de theorie beter te testen met de werkelijkheid, ons beter in staat stellen te bepalen welke theorie wanneer relevant is, en ons suggesties kunnen geven voor nieuwe, betere theorieën.

Bibliography

- [1] Antonakis, J. (2017), On doing better science: From thrill of discovery to policy implications, *Leadership Quarterly*, 28(1), 5-21.
- [2] Aghion, P., & Tirole, J. (1997), Formal and real authority in organizations, *Journal of political economy*, 105(1), 1-29.
- [3] Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *The quarterly journal of economics*, 94(4), 749-775.
- [4] Akerlof, G., & R. Kranton (2005), Identity and the Economics of Organizations, *Journal of Economic Perspectives*, 19(1), 9-32.
- [5] Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American economic review*, 62(5), 777-795.
- [6] Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, 80(3), 260.
- [7] Angrist, J.D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics*, 114(2), 533-575.
- [8] Apesteguia, J., Azmat, G., & Iriberry, N. (2012), The impact of gender composition on team performance and decision-making: Evidence from the field, *Management Science*, 58(1), 78-93.

- [9] Apicella, C. L., Demiral, E. E., & Mollerstrom, J. (2017). No Gender Difference in Willingness to Compete When Competing against Self. *American Economic Review*, 107(5), 136-140.
- [10] Aranda, C., Arellano, J., & Davila, A. (2017), Organizational learning in target setting, *Academy of Management Journal*, 60(3), 1189-1211.
- [11] Arnold, M. C., & Artz, M. (2015), Target difficulty, target flexibility, and firm performance: Evidence from business units' targets, *Accounting, organizations and society*, 40, 61-77.
- [12] Azmat, G., & Iriberri, N. (2010), The importance of relative performance feedback information: Evidence from a natural experiment using high school students, *Journal of Public Economics*, 94(7), 435-452.
- [13] Azmat, G., & Iriberri, N. (2016), The provision of relative performance feedback: An analysis of performance and satisfaction, *Journal of Economics and Management Strategy*, 25(1), 77-110.
- [14] Babcock, P., Bedard, K., Charness, G., Hartman, J., & Royer, H. (2015), Letting down the team? Evidence of social effects of team incentives, *Journal of the European Economic Association*, 13(5), 841-870.
- [15] Baker, G. (1992), Incentive Contracts and Performance Measurement, *Journal of Political Economy*, 100, 598-614.
- [16] Baker, G. (2002), Distortion and Risk in Optimal Incentive Contracts, *Journal of Human Resources*, 37(4), 728-751.
- [17] Baker, G., Gibbons, R. & Murphy, K. (1994), Subjective Performance Measures in Optimal Incentive Contracts, *Quarterly Journal of Economics*, 109, 1125-1156.
- [18] Bandiera, O., Barankay, I., & Rasul, I. (2007), Incentives for managers and inequality among workers: evidence from a firm-level experiment, *Quarterly Journal of Economics*, 122(2), 729-773.

Bibliography

- [19] Bandiera, O., Barankay, I., & Rasul, I. (2009), Social connections and incentives in the workplace: Evidence from personnel data, *Econometrica*, 77(4), 1047-1094.
- [20] Bandiera, O., Barankay, I., & Rasul, I. (2013), Team incentives: evidence from a firm-level experiment, *Journal of the European Economic Association*, 11(5), 1079-1114.
- [21] Bandiera, O., Larcinese, V., & Rasul, I. (2010). Heterogeneous class size effects: New evidence from a panel of university students. *The Economic Journal*, 120(549), 1365-1398.
- [22] Bandiera, O., Larcinese, V., & Rasul, I. (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics*, 34, 13-25.
- [23] Barankay, I. (2011). Gender differences in productivity responses to performance rankings: Evidence from a randomized workplace experiment. *Unpublished working paper, The Wharton School*.
- [24] Barankay, I. (2012), Rank Incentives: Evidence from a Randomized Workplace Experiment, *mimeo*, University of Pennsylvania.
- [25] Bertrand, M., Dufo, E., & Mullainathan, S. (2004), How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1), 249-275.
- [26] Besley, T., & Ghatak, M. (2005), Competition and Incentives with Motivated Agents, *American Economic Review*, 95(3), 616-636.
- [27] Bettinger, E. P., & Baker, R. B. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, 36(1), 3-19.
- [28] Blanes-i-Vidal, J., & Nossol, M. (2011), Tournaments without prizes: evidence from personnel records, *Management Science*, 57(10), 1721-1736.

- [29] Bol, J. (2011), The Determinants and Performance Effects of Managers' Performance Evaluation Biases, *The Accounting Review*, 82, 1549-1575.
- [30] Bol, J. C., & Lill, J. B. (2015), Performance target revisions in incentive contracts: Do information and trust reduce ratcheting and the ratchet effect? *The Accounting Review*, 90(5), 1755-1778.
- [31] Bolton, P., & Dewatripont, M. (2005). *Contract theory*. MIT press.
- [32] Bordalo, P., Coffman, K. B., Gennaioli, N., & Shleifer, A. (2016). Beliefs about gender (No. w22972). *National Bureau of Economic Research*.
- [33] Bouwens, J., & Kroos, P. (2011), Target ratcheting and effort reduction, *Journal of Accounting and Economics*, 51(1), 171-185.
- [34] Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016), Employee recognition and performance: A field experiment, *Management Science*, 62(11), 3085-3099.
- [35] Breuer, K., Nieken, P., & Sliwka, D. (2013), Social Ties and Subjective Performance Evaluations: An Empirical Investigation, *Review of managerial Science*, 7(2), 141-157.
- [36] Brookins, P., Goerg, S., & Kube, S. (2017). Self-chosen goals, incentives, and effort. *Working paper*.
- [37] Budde, J. (2007), Performance Measure Congruity and the Balanced Scorecard, *Journal of Accounting Research*, 45(3), 515-539.
- [38] Bull, C. (1987), The Existence of Self-Enforcing Implicit Contracts, *Quarterly Journal of Economics*, 102, 147-159.
- [39] Burgelman, R.A. (1994), Fading Memories: A Process Theory of Strategic Business Exit in Dynamic Environments, *Administrative Science Quarterly*, 39(1), 39-56.

Bibliography

- [40] Burgess, S., Propper, C., Ratto, M., Scholder, K., von Hinke, S., & Tominey, E. (2010), Smarter Task Assignment or Greater Effort: The Impact of Incentives on Team Performance, *Economic Journal*, 120(547), 968-989.
- [41] Carroll, G. (2015), Robustness and linear contracts, *The American Economic Review*, 105(2), 536-563.
- [42] Casas-Arce, P., & Martinez-Jerez, F. A. (2009), Relative performance compensation, contests, and dynamic incentives, *Management Science*, 55(8), 1306-1320.
- [43] Chen, H., & Lim, N. (2013), Should managers use team-based contests?, *Management Science*, 59(12), 2823-2836.
- [44] Clark, D., Gill, D., Prowse, V., & Rush, M. (2017). Using goals to motivate college students: Theory and evidence from field experiments (No. w23638). *National Bureau of Economic Research*.
- [45] Cooper, D.J., & Sutter, M. (2018), Endogenous role assignment and team performance, *International Economic Review*, 59(3), 1547-1569..
- [46] Corgnet, B., Gómez-Miñambres, J., & Hernán-Gonzalez, R. (2015). Goal Setting and Monetary Incentives: When Large Stakes Are Not Enough. *Management Science*, 61(12), 2926-2944.
- [47] Corgnet, B., Gomez-Minambres, J., & Hernan-Gonzalez, R. (2016). Goal Setting in the Principal-Agent Model: Weak Incentives for Strong Performance. *Games and Economic Behavior*, 109, 311-326.
- [48] Dal Bo, E., Finan, F. & Rossi, M.A. (2013), Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service, *Quarterly Journal of Economics*, 128(3), 1169-1218.
- [49] Dalton, P.S., Gonzalez Jimenez, V.H., & Noussair, C.N. (2015). Paying with Self-Chosen Goals: Incentives and Gender Differences. *CentER Discussion Paper*; Vol. 2015-021. Tilburg: Economics.

- [50] Datar, S., Kulp, S.C. & Lambert, R.A. (2001), Balancing Performance Measures, *Journal of Accounting Research*, 39(1), 75-92.
- [51] Delfgaauw, J., & Dur, R. (2008), Incentives and Workers' Motivation in the Public Sector, *Economic Journal*, 118(525), 171-191.
- [52] Delfgaauw, J., & Dur, R. (2010), Managerial Talent, Motivation, and Self-selection into Public Management, *Journal of Public Economics*, 94(9-10), 654-660.
- [53] Delfgaauw, J., Dur, R., Non, A., & Verbeke, W. (2014), Dynamic incentive effects of relative performance pay: a field experiment, *Labour Economics*, 28, 1-13.
- [54] Delfgaauw, J., Dur, R., Non, A., & Verbeke, W. (2015), The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment, *Journal of Labor Economics*, 33(3), 521-569.
- [55] Delfgaauw, J., Dur, R., Sol, J., & Verbeke, W. (2013), Tournament incentives in the field: Gender differences in the workplace, *Journal of Labor Economics*, 31(2), 305-326.
- [56] Dessein, W., & Santos, T. (2006), Adaptive organizations, *Journal of Political Economy*, 114(5), 956-995.
- [57] Dreber, A., Von Essen, E., & Ranehill, E. (2011). Outrunning the gender gap—boys and girls compete equally. *Experimental Economics*, 14(4), 567-582.
- [58] Dur, R., & Tichem, J. (2015), Altruism and Relational Contracts in the Workplace, *Journal of Economics and Management Strategy*, forthcoming.
- [59] Eccles, R. & Crane, D. (1988), *Doing Deals: Investment Banks at Work*. Boston: Harvard Business School Press.

Bibliography

- [60] Erev, I., Bornstein, G., & Galili, R. (1993), Constructive intergroup competition as a solution to the free rider problem: A field experiment, *Journal of Experimental Social Psychology*, 29(6), 463-478.
- [61] Feltham, G., & Xie, J. (1994), Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations, *The Accounting Review*, 69(3), 429-453.
- [62] Ferreira, D., & Sah, R.K. (2012), Who Gets To the Top? Generalists versus Specialists in Organizations, *RAND Journal of Economics*, 43(4), 577-601.
- [63] Freixas, X., Guesnerie, R., & Tirole, J. (1985), Planning under incomplete information and the ratchet effect, *The review of economic studies*, 52(2), 173-191.
- [64] Friebe, G., Heinz, M., Krüger, M., & Zubanov, N. (2017), Team incentives and performance: Evidence from a retail chain, *American Economic Review*, 107(8), 2168-2203.
- [65] Fryer, R. (2013), Teacher incentives and student achievement: Evidence from New York City public schools, *Journal of Labor Economics*, 31(2), 373-407.
- [66] Gómez-Miñambres, J. (2012). Motivation through goal setting. *Journal of Economic Psychology*, 33(6), 1223-1239.
- [67] Garicano, L. (2000), Hierarchies and the Organization of Knowledge in Production, *Journal of Political Economy*, 108(5), 874-904.
- [68] Gibbons, R. (1987), Piece-rate incentive schemes, *Journal of Labor Economics*, 5(4, Part 1), 413-429.
- [69] Gibbons, R. (2005), Four formal (izable) theories of the firm? *Journal of Economic Behavior & Organization*, 58(2), 200-245.
- [70] Gibbs, M. (1995), Incentive Compensation in a Corporate Hierarchy, *Journal of Accounting and Economics*, 19(2), 247-277.

- [71] Gibbs, M., Merchant, K.A., Van der Stede, W.A. & Vargus, M.E. (2004), Determinants and Effects of Subjectivity in Incentives, *The Accounting Review*, 79(2), 409-436.
- [72] Giebe, T., & Gurtler, O. (2012), Optimal Contracts for Lenient Supervisors, *Journal of Economic Behavior & Organization*, 81, 403-420.
- [73] Glewwe, P., Ilias, N., & Kremer, M. (2010), Teacher incentives, *American Economic Journal: Applied Economics*, 2(3), 205–227.
- [74] Gneezy, U. & Rey-Biel, P. (2014), On the relative efficiency of performance pay and noncontingent incentives, *Journal of the European Economic Association*, 12(1), 62-72.
- [75] Gneezy, U., & Rustichini, A. (2000), Pay enough or don't pay at all, *Quarterly Journal of Economics*, 115(3), 791-810.
- [76] Gneezy, U., Niederle, M., & Rustichini, A. (2003), Performance in competitive environments: Gender differences, *Quarterly Journal of Economics*, 118(3), 1049-1074.
- [77] Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics*, 118(3), 1049-1074.
- [78] Goerg, S.J., & Kube, S. (2012). Goals (th)at Work—Goals, Monetary Incentives, and Workers' Performance. *MPI Collective Goods Preprint*, (2012/19).
- [79] Goodman S.F., & Turner, L.J. (2013), The design of teacher incentive pay and educational outcomes: Evidence from the New York City bonus program, *Journal of Labor Economics*, 31(2), 409-420.
- [80] Green, C., & Heywood, J. S. (2008), Does performance pay increase job satisfaction?, *Economica*, 75(300), 710-728.

Bibliography

- [81] Grossman, S. J., & Hart, O. D. (1986), The costs and benefits of ownership: A theory of vertical and lateral integration, *Journal of Political Economy*, 94(4), 691-719.
- [82] Guth, D.G. & MacMillan, I.C. (1986), Strategy Implementation versus Middle Management Self-interest, *Strategic Management Journal*, 7, 313-327.
- [83] Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational behavior and human performance*, 16(2), 250-279.
- [84] Handy, F. & Katz, E. (1998), The Wage Differential between Nonprofit Institutions and Corporations: Getting More by Paying Less?, *Journal of Comparative Economics*, 26, 246-261.
- [85] Hart, O., & Holmström, B. (2010), A theory of firm scope, *The Quarterly Journal of Economics*, 125(2), 483-513.
- [86] Hart, O., & Moore, J. (1990), Property Rights and the Nature of the Firm, *Journal of Political Economy*, 98(6), 1119-1158.
- [87] Heywood, J. S., & Wei, X. (2006), Performance pay and job satisfaction, *Journal of Industrial Relations*, 48(4), 523-540.
- [88] Hollenbeck, J.R., Williams, C.R., & Klein, H.J. (1989). An empirical examination of the antecedents of commitment to difficult goals. *Journal of Applied Psychology*, 74(1), 18.
- [89] Holmström, B. (1979). Moral hazard and observability. *The Bell journal of economics*, 74-91.
- [90] Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 324-340.
- [91] Holmström, B. (2017). Pay for performance and beyond. *American Economic Review*, 107(7), 1753-77.

- [92] Holmström, B., and P. Milgrom (1991), Multi-Task Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, *Journal of Law, Economics and Organization*, 7, 24-52.
- [93] Holmstrom, B., & Milgrom, P. (1987), Aggregation and linearity in the provision of intertemporal incentives, *Econometrica*, 303-328.
- [94] Hsiaw, A. (2013). Goal-setting and self-control. *Journal of Economic Theory*, 148(2), 601-626.
- [95] Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature. *Journal of applied psychology*, 92(5), 1332.
- [96] Kamphorst, J. & Swank, O. (2018), The Role of Performance Appraisals in Motivating Employees, *Journal of Economics & Management Strategy*, 27(2), 251-269.
- [97] Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of political Economy*, 100(4), 801-817.
- [98] Koch, A. K., & Nafziger, J. (2016). Goals and bracketing under mental accounting. *Journal of Economic Theory*, 162, 305-351.
- [99] Koch, A., Nafziger, J., & Nielsen, H.S. (2014). Behavioral economics of education. *Journal of Economic Behavior & Organization*, 115, 3-17.
- [100] Koch, A.K., & Nafziger, J. (2011). Self-regulation through Goal Setting. *The Scandinavian Journal of Economics*, 113(1), 212-227.
- [101] Koch, A.K., Nafziger, J., Suvorov, A., & Van de Ven, J. (2014). Self-rewards and personal motivation. *European Economic Review*, 68, 151-167.
- [102] Kofman, F., & Lawarrée, J. (1996), A Prisoner's Dilemma Model of Collusion Deterrence, *Journal of Public Economics*, 59, 117-136.

Bibliography

- [103] Kosfeld, M., & Neckermann, S. (2011), Getting more work for nothing? Symbolic awards and worker performance, *American Economic Journal: Microeconomics*, 3(3), 86-99.
- [104] Kuhnen, C.M., & Tymula, A. (2012), Feedback, self-esteem, and performance in organizations, *Management Science*, 58(1), 94-113.
- [105] Laffont, J. (1990), Analysis of Hidden Gaming in a Three-Level Hierarchy, *Journal of Law, Economics & Organization*, 6(2), 301-324.
- [106] Laffont, J., & Rochet, J. (1997), Collusion in Organizations, *Scandinavian Journal of Economics*, 99(4), 485-495.
- [107] Latham, G. P., & Marshall, H. A. (1982). The effects of self-set, participatively set and assigned goals on the performance of government employees. *Personnel Psychology*, 35(2), 399-404.
- [108] Latham, G.P., & Locke, E.A. (1979). Goal setting—A motivational technique that works. *Organizational Dynamics*, 8(2), 68-80.
- [109] Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral Economics of Education. *Handbook of the Economics of Education*, 5, 1-74.
- [110] Lavy, V. (2002), Evaluating the effect of teachers' group performance incentives on pupil achievement, *Journal of Political Economy*, 110(6), 1286-1317.
- [111] Lazear, E.P. (1986), Salaries and piece rates, *Journal of Business*, 405-431.
- [112] Levin, J. (2003). Relational incentive contracts, *American Economic Review*, 93(3), 835-857.
- [113] Levitt, S.D., List, J.A., Neckermann, S., & Sadoff, S. The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance. *American Economic Journal: Economic Policy*, 8(4), 183-219..
- [114] Li, D. (2013), Expertise vs. Bias in Evaluation: Evidence from the NIH, mimeo, Northwestern University.

- [115] Linnenbrink, E.A. (2005). The Dilemma of Performance-Approach Goals: The Use of Multiple Goal Contexts to Promote Students' Motivation and Learning. *Journal of Educational Psychology*, 97(2), 197-213.
- [116] List, J., Sadoff, S., & Wagner, M. (2011), So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design, *Experimental Economics*, 14, 439-457.
- [117] Locke, E.A. (1996). Motivation through conscious goal setting. *Applied and Preventive Psychology*, 5(2), 117-124.
- [118] Locke, E.A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705-717.
- [119] Locke, E.A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science*, 15(5), 265-268.
- [120] Malcomson, J. M. (1984), Relational Incentive Contracts, in R. Gibbons and J. Roberts (eds.), *The Handbook of Organizational Economics*. Princeton, NJ: Princeton University Press.
- [121] Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1), 112-45.
- [122] McCausland, W. D., Pouliakas, K., & Theodossiou, I. (2005), Some are punished and some are rewarded: a study of the impact of performance pay on job satisfaction, *International Journal of Manpower*, 26(7/8), 636-659.
- [123] Mookherjee, D., (2013), Incentives in Hierarchies, in R. Gibbons and J. Roberts (eds.), *The Handbook of Organizational Economics*. Princeton, NJ: Princeton University Press.
- [124] Morisano, D., Hirsh, J. B., Peterson, J. B., Pihl, R. O., & Shore, B. M. (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. *Journal of Applied Psychology*, 95(2), 255-264.

Bibliography

- [125] Muehlheusser, G., Schneemann, S., & Sliwka, D. (2016), The impact of managerial change on performance: The role of team heterogeneity, *Economic Inquiry*, 54(2), 1128-1149.
- [126] Muralidharan, K., & Sundararaman, V. (2011), Teacher performance pay: Experimental evidence from India, *Journal of Political Economy*, 119(1), 39-77.
- [127] Niederle, M., & Vesterlund, L. (2011). Gender and Competition. *Annual Review of Economics*, 3(1), 601-630.
- [128] Ordóñez, L.D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009). Goals gone wild: The systematic side effects of overprescribing goal setting. *The Academy of Management Perspectives*, 23(1), 6-16.
- [129] Pearce, D.G., & Stacchetti, E. (1998), The Interaction of Implicit and Explicit Contracts in Repeated Agency, *Games and Economic Behavior*, 23, 75-96.
- [130] Petrescu, A. I., & Simmons, R. (2008), Human resource management practices and workers' job satisfaction, *International Journal of Manpower*, 29(7), 651-667.
- [131] Prasad, S., (2009), Task Assignment and Incentives: Generalists versus Specialists, *Rand Journal of Economics*, 40(2), 380-403.
- [132] Prendergast, C. (2002), The tenuous trade-off between risk and incentives, *Journal of political Economy*, 110(5), 1071-1102.
- [133] Prendergast, C. (2007), The Motivation and Bias of Bureaucrats, *American Economic Review*, 97(1), 180-196.
- [134] Prendergast, C., & Topel, R. (1996), Favoritism in Organizations, *Journal of Political Economy*, 104(5), 958-978.

- [135] Raith, M. (2008), Specific Knowledge and performance measurement, *RAND Journal of Economics*, 39(4), 1059-1079.
- [136] Sackett, A.M., Wu, G., White, R.J., & Markle, A. (2014). Harnessing optimism: How eliciting goals improves performance. *Working Paper*.
- [137] Sanders, M., & Chonaire, A.N. (2015). "Powered to Detect Small Effect Sizes": You keep saying that. I do not think it means what you think it means (No. 15/337). *Working Paper Department of Economics, University of Bristol, UK*.
- [138] Schippers, M.C., Scheepers, A.W., & Peterson, J.B. (2015). A scalable goal-setting intervention closes both the gender and ethnic minority achievement gap. *Palgrave Communications*, 1.
- [139] Schmidt, K.M., & M. Schnitzer (1995), The Interaction of Explicit and Implicit Contracts, *Economic Letters*, 48(2), 193-199.
- [140] Schunk, D.H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, 25(1), 71-86.
- [141] Seijts, G.H., Latham, G.P., Tasa, K., & Latham, B.W. (2004). Goal setting and goal orientation: An integration of two different yet related literatures. *Academy of Management Journal*, 47(2), 227-239.
- [142] Shavell, S. (1979). Risk sharing and incentives in the principal and agent relationship. *The Bell Journal of Economics*, 55-73.
- [143] Simon, H. A. (1951), A formal theory of the employment relationship, *Econometrica*, 293-305.
- [144] Smithers, S. (2015). Goals, motivation and gender. *Economics Letters*, 131, 75-77.
- [145] Springer, M.G., Ballou, D., Hamilton, L., Le, V.-N., Lockwood, J.R., McCaffrey, D.F., Pepper, M., & Stecher, B.M. (2010), Teacher pay for perfor-

Bibliography

- mance: Experimental evidence from the project on incentives in teaching, Nashville, TN: National Center on Performance Incentives at Vanderbilt University.
- [146] Suvorov, A. & van de Ven, J. (2009), Discretionary Rewards as a Feedback Mechanism, *Games and Economic Behavior*, 67, 665-681.
- [147] Suvorov, A., & Van de Ven, J. (2008). Goal setting as a self-regulation mechanism. *Working Paper*.
- [148] Thiele, V. (2013), Subjective Performance Evaluations, Collusion, and Organizational Design, *Journal of Law, Economics, & Organization*, 29(1), 35-59.
- [149] Tirole, J. (1986), Hierarchies and Bureaucracies: On the Role of Collusion in Organizations, *Journal of Law, Economics, and Organization*, 2, 181-214.
- [150] Travers, C.J., Morisano, D., & Locke, E.A. (2015). Self-reflection, growth goals, and academic outcomes: A qualitative study. *British Journal of Educational Psychology*, 85(2), 224-241.
- [151] Vafaï, K. (2002), Preventing Abuse of Authority in Hierarchies, *International Journal of Industrial Organization*, 20, 1143-1166.
- [152] Vafaï, K. (2010), Opportunism in Organizations, *Journal of Law, Economics, & Organization*, 26(1), 158-181.
- [153] Van Lent, M. (2017). Goal Setting under Uncertainty: Rigid vs. Flexible Goals. *Working Paper*.
- [154] Weitzman, M. L. (1980), The "ratchet principle" and performance incentives, *The Bell Journal of Economics*, 302-308.
- [155] Welbournel, T. M., & Cable, D. M. (1995), Group incentives and pay satisfaction: Understanding the relationship through an identity theory perspective, *Human Relations*, 48(6), 711-726.

- [156] Woods, A. (2012), Subjective Adjustments to Objective Performance Measures: The Influence of Prior Performance, *Accounting, Organizations and Society*, 37, 403-425.
- [157] Zabochnik, J. (2014), Subjective Evaluations with Performance Feedback, *RAND Journal of Economics*, 45(2), 341-369.
- [158] Zehnder, C., Herz, H., & Bonardi, J.-P. (2017), A productive clash of cultures: Injecting economics into leadership research, *Leadership Quarterly*, 28(1), 65-85.

Bibliography

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University 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:

- 696. H. ZHANG, *Dealing with Health and Health Care System Challenges in China: Assessing Health Determinants and Health Care Reforms*
- 697. M. VAN LENT, *Essays on Intrinsic Motivation of Students and Workers*
- 698. R.W. POLDERMANS, *Accuracy of Method of Moments Based Inference*
- 699. J.E. LUSTENHOUWER, *Monetary and Fiscal Policy under Bounded Rationality and Heterogeneous Expectations*
- 700. W. HUANG, *Trading and Clearing in Modern Times*
- 701. N. DE GROOT, *Evaluating Labor Market Policy in the Netherlands*
- 702. R.E.F. VAN MAURIK, *The Economics of Pension Reforms*
- 703. I. AYDOGAN, *Decisions from Experience and from Description: Beliefs and Probability Weighting*
- 704. T.B. CHILD, *Political Economy of Development, Conflict, and Business Networks*
- 705. O. HERLEM, *Three Stories on Influence*
- 706. J.D. ZHENG, *Social Identity and Social Preferences: An Empirical Exploration*
- 707. B.A. LOERAKKER, *On the Role of Bonding, Emotional Leadership, and Partner Choice in Games of Cooperation and Conflict*
- 708. L. ZIEGLER, *Social Networks, Marital Sorting and Job Matching. Three Essays in Labor Economics*
- 709. M.O. HOYER, *Social Preferences and Emotions in Repeated Interactions*
- 710. N. GHEBRIHIWET, *Multinational Firms, Technology Transfer, and FDI Policy*
- 711. H.FANG, *Multivariate Density Forecast Evaluation and Nonparametric Granger Causality Testing*
- 712. Y. KANTOR, *Urban Form and the Labor Market*
- 713. R.M. TEULINGS, *Untangling Gravity*
- 714. K.J.VAN WILGENBURG, *Beliefs, Preferences and Health Insurance Behavior*
- 715. L. SWART, *Less Now or More Later? Essays on the Measurement of Time Preferences in Economic Experiments*
- 716. D. NIBBERING, *The Gains from Dimensionality*
- 717. V. HOORNWEG, *A Tradeoff in Econometrics*

718. S. KUCINSKAS, *Essays in Financial Economics*
719. O. FURTUNA, *Fiscal Austerity and Risk Sharing in Advanced Economies*
720. E. JAKUCIONYTE, *The Macroeconomic Consequences of Carry Trade Gone Wrong and Borrower Protection*
721. M. LI, *Essays on Time Series Models with Unobserved Components and Their Applications*
722. N. CIURILĂ, *Risk Sharing Properties and Labor Supply Disincentives of Pay-As-You-Go Pension Systems*
723. N.M. BOSCH, *Empirical Studies on Tax Incentives and Labour Market Behaviour*
724. S.D. JAGAU, *Listen to the Sirens: Understanding Psychological Mechanisms with Theory and Experimental Tests*
725. S. ALBRECHT, *Empirical Studies in Labour and Migration Economics*
726. Y.ZHU, *On the Effects of CEO Compensation*
727. S. XIA, *Essays on Markets for CEOs and Financial Analysts*
728. I. SAKALAUSKAITE, *Essays on Malpractice in Finance*
729. M.M. GARDBERG, *Financial Integration and Global Imbalances.*
730. U. THÜMMEL, *Of Machines and Men: Optimal Redistributive Policies under Technological Change*
731. B.J.L. KEIJERS, *Essays in Applied Time Series Analysis*
732. G. CIMINELLI, *Essays on Macroeconomic Policies after the Crisis*
733. Z.M. LI, *Econometric Analysis of High-frequency Market Microstructure*
734. C.M. OOSTERVEEN, *Education Design Matters*
735. S.C. BARENDSE, *In and Outside the Tails: Making and Evaluating Forecasts*
736. S. SÓVÁGÓ, *Where to Go Next? Essays on the Economics of School Choice*
737. M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*
738. M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*
739. R.J. DÖTTLING, *Essays in Financial Economics*
740. E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*
741. Y.M. KUTLUAY, *The Value of (Avoiding) Malaria*
742. A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*
743. B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*
744. R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*
745. P.M.G. GOLEC, *Essays in Financial Economics*