Compensation, Favoritism, and Adverse Selection - Essays on Managerial Incentives in Firms

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2012

vorgelegt von

Diplom-Volkswirt Claus Ortwin Herbertz

aus

Köln

Referent:	Professor Dr. Dirk Sliwka
Korreferent:	Professor Dr. Bernd Irlenbusch
Tag der Promotion:	03.02.2012

To Christina and Theodor

Preface

This dissertation came into being in the autumn of 2007, as I entered the Graduate School for Risk Management at the University of Cologne. At the same time, the Deutsche Forschungsgemeinschaft kindly granted me three years of generous support for my dissertation project. Let me first and foremost thank my primary dissertation supervisor, Dirk Sliwka, for his great personal, academic and professional support. Over the last 4 years, he supported good ideas, revealed bad ones and constantly gave me input by pointing out subtle economic points or summarizing a bunch of papers in a sentence. Thank you Dirk!

Thanks also to my co-authors Johannes Berger, and Torsten Biemann for their great work, constant input of ideas and relentless criticisms. I am not exaggerating when I say that they substantially contributed to my thesis by their stimulating ideas and countless discussions.

I was very lucky to meet my colleagues Christof Wiechers, Frowin Schulz, Kathrin Breuer, Martin Ruppert, Miriam Breunsbach and Tobias Wickern at the Graduate School of Risk Management. During the last 4 years, I spent more time with them than with anyone else. Every single one of them taught me something and made my work more successful and my life more enjoyable. Thanks to Christof for wasting evenings with music and thanks to Frowin for his impeccable conscientiousness and discipline. Thanks to Kathrin for unconditioned happiness and thanks to Miriam for taking care. Thanks to Tobias for his cool temper and modesty and thanks to Martin for teaching me to write down the problem first. I also want to mention people from the 2008 Graduate School class: Julius Schnieders, Stefan Erdorf, Konstantin Glombek, Pavlo Bazovkin, Alwin Stöter and Nico Heinrichs. They enriched my scientific landscape and made life in Graduate School much more vivid.

I also want to thank my colleagues at the chair of Personnel Economics, Anastasia Danilov, Patrick Kampkötter and Nannan Zhou simply for being helpful, inspiring and, if needed, patient. It was a great time with you! Thanks to Björn Hartig and Felix Kölle for their superb work in joint teaching projects!

Thank you Alexander Creson, Behnud Djawadi, David Hirschfeld, Christian Ruppert, Isabella Cetnarowski, Julia Sohn, Katharina Laske, Michael Lüttjohann, Philip Aminski, Thorsten Hormesch, Timo Vogelsang, Tobias Hinze and Ursula Schuh for your outstanding research assistance, IT-support and getting things done. I thank Bernd Irlenbusch, Gari Walkowitz, Julian Conrads, Miriam Mezger and Rainer Rilke for tough discussions and fruitful suggestions during our brown bag seminar. A special thanks goes to Bernd for being my second dissertation supervisor!

I thank the Great Place to Work Institute, namely Frank Hauser, Johanna Kirsch and Karsten Schulte, for an uncomplicated and successful research collaboration.

My friends Julian and Zsófi Orlishausen, Joachim and Nicole Horvath, Thomas Böttinger, Oliver Hennies and my family always provided me with the encouragement, support and care that I needed.

Last, but not least, I thank my wife Christina for her love, trust, patience and care and my son Theodor, who taught me that even a doctoral thesis is only one thing among many.

Contents

1 Introduction **1**

2 Managerial Incentives and Favoritism in Promotion Decisions - Theory and Field Evidence **7**

2.1 Introduction 7

2.2 Related Literature 9

2.3 The Model 11

2.4 Equilibrium Analysis 12

2.5 Data and Hypotheses 15

2.6 Results 18

2.6.1 Managerial Incentives and Favoritism 18

2.6.2 Robustness and Potentially Omitted Variables 26

2.7 Conclusion 29

2.8 Appendix to Chapter 2 30

3 When Higher Prizes Lead to Lower Efforts - The Impact of Favoritism in Tournaments **35**

3.1 Introduction 35

3.2 The Model 36

3.3 Equilibrium Analysis 37

3.4 Conclusion 41

3.5 Appendix to Chapter 3 42

4 Social Ties, Incentives and Adverse Selection **44**

4.1 Introduction 44

4.2 Design and Hypotheses 47
4.3 Results 51
4.4 Conclusion 54
4.5 Appendix to Chapter 4 56
4.5.1 Tables and Figures 56
4.5.2 Instructions 58

5 Result-Based Sampling in Experimental Economics 62
5.1 Introduction 62
5.2 Type I Error Inflation in Statistical Hypothesis Testing 64
5.3 Simulation and Results 66
5.4 Determinants of Type I Error Inflation 68
5.5 Discussion 71
5.6 Power Analysis 72
5.7 Conclusion 74
5.8 Appendix to Chapter 5 75

Bibliography 75

List of Tables

2.1 The Relationship between Managerial Incentives and Favoritism 21
2.2 Ordered Probit Marginal Effects 22
2.3 The Relationship between Managerial Incentives and Favoritism - Quadratic and Nonparametric Models 25
2.4 The Relationship between Managerial Incentives and Favoritism - Controls I 27
2.5 Descriptive Statistics of Variables on Firm Level 31
2.6 Descriptive Statistics of Variables on Individual Level 32
2.7 The Relationship between Managerial Incentives and Favoritism - OLS Analysis 33
2.8 The Relationship between Managerial Incentives and Favoritism - Controls II 34

4.1 2x2 Treatment Design 49
4.2 Decision Player A 51
4.3 Decision Player C 52
4.4 Decision Player A: p-values for a two sided (above diagonal) and one sided (below diagonal) Fisher exact Test between Treatments 56
4.5 Decision Player C: p-values for a two sided (above diagonal) and one sided (below diagonal) Fisher exact Test between Treatments 56

5.1 Simulated Type I Errors from Figure 5.3 75
5.2 Simulated Type I Errors from Figure 5.4 75

List of Figures

2.1 Probability that more able agents are promoted as a function of β. Graphs for $\Delta\eta = 0.1$ (solid), $\Delta\eta = 0.2$ (dashed) and $\Delta\eta = 0.4$ (dotted). 15
2.2 Utilization of Performance related Pay across German Industries 17
2.3 Histogram of Proportions of positive Answers in Firms with and without Manager PRP 19
2.4 Histogram of Proportions of positive Answers in Firms with and without Manager Gain Sharing Plans 19

4.1 A 2-Stage Pie Sharing Game 48
4.2 The empirical Distribution of the Friendship Measure 57
4.3 The empirical Distribution of Within-Group Differences in Friendship Measures 57

5.1 The Simulation Process 67
5.2 Histogram of p-values after first and second Sampling 69
5.3 α_{RBS} in Dependence of the Sampling Threshold p_{RBS} and the Number of Samplings k 70
5.4 α_{RBS} in Dependence of the Sampling Threshold p_{RBS} and the Number of Samplings k for a Mann-Whitney U Test 76

Chapter 1

Introduction

In this thesis, we investigate how monetary incentive schemes influence promotion and distribution decisions in the presence of favoritism.

In chapter 2, we theoretically analyze the relationship between managerial incentives and promotion quality in the presence of favoritism, stating that incentives crowd out favoritism and lead to better promotion decisions. Testing this hypothesis empirically, we find a positive relationship between the use of managerial incentives and promotion quality in German firms. In chapter 3, we point out a drawback of high promotion prizes in tournaments with favoritism: In the presence of favoritism supervisors gain utility by awarding the tournament prize to their favored agent and are less likely to promote the more able agent. For large prizes, this effect outweighs the incentive effect of the tournament prize. Consequently, the agents' effort declines in the prize. In chapter 4, we experimentally investigate favoritism-induced selection effects by forming 3-person groups with two friends and an anonymous player in a distribution game. Anonymous players avoid the distribution game, fearing harmful collaborations of the befriended participants. Incentives for decision-makers partially crowd out favoritism and the anonymous player enters the distribution game. Chapter 5 deals with a methodological problem in experimental economics. It is a common practice to conduct experimental sessions, evaluate the resulting data and conduct further sessions if no significant results are attained. We illustrate that this approach leads to a Type I Error inflation and make suggestions for better experimental planning.

One core problem in contract theory is the alignment of goals in a principal agent setting. The principal wants the agent to supply effort, whereas the agent does not want to exert costly and unobservable effort once a contract is signed. Proper incentives mitigate the moral hazard problem by tying the agent's compensation to an observable signal which consists of her unobservable effort and a noise component.[1] One key assumption is that a proper signal is contractible. Typically, this assumption is fulfilled in the field if the agent's output is quantifiable in terms of pieces produced, revenue raised, cars repaired or trees planted. Not surprisingly, in this environment, incentive schemes do a pretty good job in increasing worker productivity when compared to fixed wages (Lazear (2000) and Shearer (2004)). But especially white-collar jobs often involve tasks which are difficult to measure and make an objective performance evaluation too costly, if not impossible (Murphy and Cleveland (1995) and Prendergast (1999)). In this case, the pay for performance approach is not feasible in the simple form described above. This problem becomes even worse when an agent works on multiple tasks. For applying the optimal incentive scheme, not only one, but multiple signals are needed. If one or more signals are not feasible, the multi tasking problem applies and incentivizing a subset of tasks leads to the disregard of unincentivized tasks (Holmström and Milgrom (1991)).

One solution to this problem is the use of subjective performance evaluations by supervisors. Since a supervisor has a picture of an agent's tasks and performance, her subjective evaluation should be a decent performance signal. Though a fair subjective evaluation is not contractible, the agent might accept contracts involving subjective evaluations due to repeated interaction or the supervisor's reputation or trustworthiness. Despite the fact that several findings point out the advantages of subjective performance evaluations (Baker et al. (1994) and Baiman and Rajan (1995)) there is a major drawback: Supervisors have discretion in evaluating an agent's performance. If

[1] See Bolton and Dewatripont (2005) for an overview. For seminal contributions see f.i. Jensen (1976), Grossman and Hart (1983) and Holmström and Milgrom (1987).

a supervisor is not the residual claimant, but is herself an agent, she might follow her own interests when evaluating subordinates, rather than reporting her true expectations about the agent's performance. Indeed, a couple of studies have pointed out political reasons as the primary determinant for subjective evaluations (Longenecker et al. (1987) and Bjerke et al. (1987)). According to anecdotal evidence in Ittner et al. (2003), discretion in balanced scorecard systems leads to extreme forms of favoritism and arbitrariness by supervising managers.

Prendergast and Topel (1996) theoretically analyze this setting by introducing favoritism in a LEN framework. Technically speaking, the supervisor gains utility from the well-being of her subordinates. This may lead to low-powered incentives for the subordinates, since higher powered incentives also strongly affect the supervisor's utility, and thus, lead to higher distortions in the evaluations.

In chapter 2, we apply the approach from Prendergast and Topel (1996) on job promotion tournaments. If a manager has preferences over agents, i.e., likes agent A more than agent B, she has a natural inclination not to promote the best, but the best liked agent. In the absence of managerial incentives, the manager will always promote her favorite agent. However, in the presence of incentives, the manager has to trade off between utility from favoritism and utility from promoting the best performing agent. The likelihood of promoting the better agent increases with the incentives for the promoting manager. This theoretical result suggests a real world correlation between managerial incentives and promotion quality in firms. Furthermore, the model suggests that in the presence of weak or moderate favoritism, low managerial incentives already have a substantial impact on promotion quality. Medium or high managerial incentives only slightly increase the promotion quality compared to low incentives. Hence, the model predicts similar correlations for promotion quality with low, medium and high managerial incentives.

We empirically test these predictions with a representative employer-employee matched database containing 36,000 employee survey observations from 305 firms. In addition to the survey, the dataset contains firm-level information about compensation and HR practices in the respective firms. We measure promotion quality by employees' degree of agreement to the statement "*Promotions go to those who best deserve them*" on a 5-point Likert scale. As explanatory variables, we use the existence and strength of managerial performance related pay and the utilization of gain sharing plans. The ordered probit regression reveals that in firms using performance related pay (gain sharing plans) for their managers, employees are 8.3% (8.8%) more likely to have a positive opinion about their firm's promotion policy. To test the second theoretical prediction, we regress promotion quality on 3 dummy variables, which capture whether the variable pay component in a firm makes up 1-10%, 10-20%, or more than 20% of a manager's total compensation. As theoretically suggested, firms which pay 10-20% or more than 20% of their managers' compensation by variable components do not have a higher perceived promotion quality compared to firms which only pay 1-10% in variable terms.

In chapter 3, we use a theoretical setup similar to that of chapter 2 to analyze the impact of favoritism on the agent's effort supply in promotion tournaments. As we already show in chapter 2, favoritism distorts the promotion decision to the favor of the more liked agent and reduces the winning chances of the less liked agent. Lower winning chances lead to lower effort supply by the less liked agent. The favored agent anticipates her competitor's rationale and also lowers her effort. This effect is already known from tournaments with favorites and underdogs. Once a player has a winning margin, the underdog reduces effort, and consequently, the favorite reduces her effort also.[2] Different to the standard case, the winning margin is not constant, but is increasing in the tournament prize if favoritism is in place. Since the supervisor gains from the favored agent's utility, high tournament prizes lead to a stronger promotion distortion and an even bigger advantage for the favored agent. The higher the prize, the higher the distortion by the supervisor, and the lower is the winning chances for the less liked agent.

[2]Note that from the experimental point of view, underdogs regulary supply substantially more effort than theoretically predicted. (Bull et al. (1987), Schotter and Weigelt (1992) and van Dijk et al. (2001))

If this effect is stronger than the pure incentive effect of the higher prize, agents' effort decrease in the tournament prize. Our results suggest, that in the presence of favoritism and subjective performance evaluations, high tournament prizes might backfire with low effort supply.

The models described in chapter 2 and chapter 3 abstract from agents' participation constraints. Disfavored agents have no possibility of leaving the tournament or the company. In a wider approach with strong favoritism or low job search costs, agents might react to favoritism by quitting jobs. An example for strong favoritism or even primogeniture are family firms in which family members are often promoted quicker and receive higher wages than non-family members (Pérez-González (2006)). If promotions are quicker due to family affiliation, non-family members have lower chances in promotion tournaments and might leave the company. However, incentivizing efficiency-centered decisions might reduce favoritism in these organizations and encourage non-favored agents to enter, or not to leave the company. In chapter 4, we follow this idea in the laboratory. We use a 2-stage 3-player game in which an unrelated anonymous Player A has to decide whether to enter a distribution game or to take an outside option. If she enters the game her payoff depends on the goodwill of Player C, who distributes money between Player A and Player B. If Player C and Player B are friends and Player C receives a fixed wage for her decision, Player A refrains from entering the game because of favoritism. She is right in doing so, since Player C strongly favors her friend in the distribution decision. However, incentivizing Player C for choosing an efficient allocation crowds out favoritism and Player A is more inclined to enter the game. The results suggest that unrelated agents anticipate the incentive structure in organizations and expect efficiency enhancing incentives to reduce favoristic behavior.

Chapter 5 deals with a practice in experimental economics which we name result-based sampling (RBS). Experimenters often conduct experiments in period 1, explore their data and decide upon on the basis of their findings how to proceed. If they find significant results, they often finish the experiment and aim for publication. If the data does not show up the suggested results, they abort the project. However, if results are close to significance, many researchers increase their sample size by conducting further sessions in period 2. The guiding principle behind this approach is that the treatment effect in question might be too small for detection with the current sample size, but may be detected with an increased sample. This argument is valid at first glance, since increasing the sample size indeed reduces the likelihood of not rejecting the null hypothesis when the null hypothesis is indeed false. But RBS has a crucial drawback: Given the null hypothesis to be true, 5% of all researchers come out with wrong positive results after period 1. Some of the remaining 95% of researchers apply RBS, increase their sample size in period 2 and gain significant results. Summing up significant results from the first and the second period shows that substantially more than 5% of all projects end up with wrong positive results. Under reasonable assumptions and a significance level of 5%, RBS leads to a Type I Error of 7-8%. To overcome RBS, we propose to determine a target sample size before conducting experiments. If the target sample size is reached the experimenter should not further increase the sample size, because this would lead to a Type I Error inflation.

Chapter 2

Managerial Incentives and Favoritism in Promotion Decisions - Theory and Field Evidence[1]

2.1 Introduction

In most jobs work performance is not perfectly reflected in objective performance measures. As a consequence superiors are often asked to rate their subordinates subjectively, which gives them the possibility to favor one subordinate over another. As a consequence, performance appraisals may be biased, not reflecting actual work performance but the supervisor's personal preferences for her subordinates. If internal promotion decisions are based on subjective performance appraisals, favoritism may eventually result in not promoting the best but those who are best liked.

To avoid favoritism, pay and promotions are sometimes solely determined by seniority and other bureaucratic rules which of course also bear the risk of poor promotion decisions (Prendergast and Topel (1996), Marsh (1960)). Another solution has been suggested by Prendergast and Topel (1993) who

[1]This chapter is based upon Berger et al. (2011).

claim that "*a means of aligning the supervisor's incentives with those of the organization is to tie rewards to promotion and to make the supervisor responsible for the output of the job to which his subordinates are promoted*" (Prendergast and Topel (1993), p. 360).

In this paper we investigate this idea in detail and analyze the relationship between managerial incentives and promotion decisions. First, we theoretically explore the effect of favoritism on managers' promotion decisions. Favoritism indeed leads to poor promotion decisions as the more able subordinate is less likely promoted. Workers anticipate their superior's bias and reduce efforts in promotion tournaments. Thus favoritism not only harms the company ex-post by putting the wrong people into management positions but it also reduces effort supply and therefore expected company earnings ex-ante. Tying the manager's salary to workers' performances makes favoritism costly and leads to better promotion decisions.

To test the theoretical predictions, we use a unique employer-employee matched data set collected on behalf of the German Federal Ministry of Labor and Social Affairs by the Great Place to Work Institute, a company specialized in conducting employee surveys, in 2006. The dataset is a representative sample of 305 German firms containing company-level information on management practices. In addition, in each of the firms an employee-survey has been conducted, containing detailed information about employee perceptions of approximately 36,000 individuals. In contrast to researchers who try to assess promotion quality as an outsider to the firm, employees have inside knowledge and are in a good position to judge whether indeed the best people are promoted. We thus focus on a standardized item contained in the employee-survey which measures whether "*promotions go to those who best deserve them*".

Indeed, we find that the quality of promotion decisions is significantly higher in firms which use managerial incentive schemes. Furthermore, the data suggest that even lower powered incentive schemes suffice to reduce favoritism in promotions which is in line with the theoretical prediction.

The remainder of the paper is organized as follows. In the next section we summarize the related literature. In sections 2.3 and 2.4 we theoretically derive predictions on the consequences of managerial pay for promotion decisions. Section 2.5 describes our dataset and the main variables used for the empirical analysis. Section 2.6 includes the main empirical results while section 2.7 discusses obtained results and summarizes our main findings.

2.2 Related Literature

The role of favoritism and social connectivity in firms has gained considerable attention in theoretical economic research in recent years. Prendergast and Topel (1996) were among the first who analyze how personal preferences towards employees may lead to favoritism and biased performance appraisals. In their model supervisors derive utility from biasing performance appraisals according to individual preferences. Favoritism then leads to a misallocation of workers to jobs and higher powered worker incentives aggravate this problem. Prendergast (2002) extends this framework and shows that noisy environments reduce favoritism-induced distortions. Fairburn and Malcomson (2001) illustrate that bribery-induced favoritism offsets the effect of monetary bonus schemes and suggest job promotion tournaments to mitigate the incentives to bribe and thus reduce favoritism.

A couple of studies empirically investigate the harmful influence of favoritism. Longenecker et al. (1987) and Bjerke et al. (1987), for instance, examine determinants of performance evaluations in a US company and the US Navy respectively. Both studies claim that political considerations rather than true performance are reflected in subjective evaluations. This is especially true if performance appraisals are tied to bonuses. Ittner et al. (2003) analyze a balanced scorecard bonus plan which is based on supervisors' subjective evaluations. Even if financial measures for evaluating subordinates are available, supervisors' discretion leads to strong favoritism in employees' bonus payments in the studied company and finally to the abolishment of the scorecard. Breuer et al. (2010) analyze personnel data from a call center organization arguing that social ties triggered by repeated interaction or small team size lead to biased performance evaluations by supervisors.

Several other studies have examined restaffing decisions in the presence of family ties, which can be seen as a prominent case of favoritism in firms. Pérez-González (2006) report a faster career as well as higher wages for family members in family firms. Kramarz and Skans (2007) find that young Swedish men frequently work in their father's plant while having higher initial wages and worse school grades than comparable colleagues. Bennedsen et al. (2007), Pérez-González (2006) and Vilallonga and Amit (2006) also find that CEO family succession leads to a significant drop in family firm performance displaying the inefficiencies caused by favoritism in succession and promotion politics.[2]

While the existence and negative influence of favoritism is well documented, possible remedies for it are less intensely studied. One important exception is Bandiera et al. (2009) who analyze an exogenous change from a fixed wage to a bonus scheme for supervisors in a large agricultural company. Under fixed wages managers favor socially connected workers by granting them a stronger support which leads to a large productivity gap between socially connected and socially unconnected workers. When supervisors receive a bonus based on workers' output they reallocate their support towards high ability workers, causing a significant overall increase in productivity. In this paper we show that higher powered managerial incentive pay also substantially affects the quality of promotion decisions and we provide empirical evidence based on a large and representative sample of firms.

Finally, by combining survey data on management practices with more objective information across larger samples on different firms, our paper adds to the emerging literature on investigating key issues in personnel economics and the economics of organizations as for instance recently advocated in (Bloom and Van Reenen, 2007, 2010).

[2] Two exceptions to these findings are Sraer and Thesmar (2007) and Anderson and Reeb (2003) who report a positive correlation between performance and the presence of family CEOs.

2.3 The Model

We consider a 3-stage model with a top manger M and two heterogeneous agents $i = A, B$ competing for a middle manger position. In the first period, agents choose an unobservable effort level e_{i1} and produce outputs

$$s_{i1} = a_i + e_{i1} + \varepsilon_{i1} \tag{2.1}$$

where $a_i \sim N\left(m_a, \sigma_a^2/2\right)$ denotes agent i's time invariant and unknown ability. The error term is also normally distributed with $\varepsilon_{i1} \sim N\left(0, \sigma_\varepsilon^2/2\right)$. We assume a_i and ε_{i1} to be independent and their distributions to be common knowledge. Providing effort yields effort costs $\frac{c}{2}e_i^2$. All players are risk neutral. Period 1 profit is given by

$$\Pi_1 = s_{A1} + s_{B1}.$$

In period 2 top manager M observes performance signals s_{i1} and chooses which agent $\phi \in \{A, B\}$ is to be promoted to the middle manager position. The promoted agent receives a wage increase Δw.

In period 3, agents choose their effort level e_{i3}, again produce $s_{i3} = a_i + e_{i3} + \varepsilon_{i3}$ and generate company profit

$$\Pi_3 = k \cdot s_{\phi 3} + s_{-\phi 3} \tag{2.2}$$

where $s_{\phi 3}$ and $s_{-\phi 3}$ are the outputs of the promoted and non-promoted agent. With $k > 1$, we assume a middle manager's performance to have larger impact on company profit as the non-promoted agent.

The agents' utilities are simply the sum of their expected wages minus their effort costs. They receive no fix wage component. M's wage is given by $\alpha + \beta \cdot (\Pi_1 + \Pi_3)$ where α is a fixed wage and β measures the extent of profit sharing. Finally, we assume that the top manager may personally like the two agents to a different extent and therefore favor one over the other. Similar to Prendergast and Topel (1996) and Prendergast (2002) she receives a utility from favoritism of $\eta_\phi \cdot \Delta w$ such that the wage increase Δw awarded to the promoted agent is weighted with a preference parameter η_i. Hence, her overall utility is

$$\alpha + \beta \cdot (\Pi_1 + \Pi_3) + \eta_\phi \cdot \Delta w.$$

We assume M's discount factor to be 1 and η_i to be known by all players. Furthermore, we take all compensation parameters with the exception of β as given and focus on the connection between β and promotion decisions.

2.4 Equilibrium Analysis

We now determine the Perfect Bayesian Equilibrium. In the absence of incentives, agents choose zero effort in stage 3 and produce according to their ability (see equation 2.2). In stage 2 top manager M's conditional expected utility for $\phi \in \{A, B\}$ depends on the agents' period 1 performance signals s_{A1} and s_{B1} and is given by

$$\begin{aligned} V_M(\phi, s_{A1}, s_{B1}) &= E\left[\alpha + \beta \cdot \Pi_3 + \eta_\phi \cdot \Delta w \mid s_{A1}, s_{B1}\right] \\ &= \alpha + \beta k \cdot E\left[a_\phi \mid s_{\phi 1}\right] + \beta \cdot E\left[a_{-\phi} \mid s_{-\phi 1}\right] + \eta_\phi \cdot \Delta w. \end{aligned}$$

Hence, M promotes agent A if $V_M(A, s_{A1}, s_{B1}) \geq V_M(B, s_{A1}, s_{B1})$ or

$$E\left[a_A \mid s_{A1}\right] - E\left[a_B \mid s_{B1}\right] \geq \frac{\Delta\eta \cdot \Delta w}{\beta(k-1)} \tag{2.3}$$

where $\Delta\eta = \eta_B - \eta_A$. In the absence of favoritism ($\Delta\eta = 0$) the rhs in equation 2.3 is zero. In this case, M's decision is solely driven by her expectations about agents' abilities and the agent who is expected to be more able gets promoted. The model then basically boils down to a standard Lazear and Rosen (1981) type tournament model. Furthermore, pure ability based promotion decisions maximize the company's third period profits in equation 2.2. If, however, favoritism matters, M gains additional utility from promoting the favored agent. The more M favors an agent, the more likely her promotion decision will not coincide with the profit maximizing decision.

Furthermore, the higher β the smaller this distortion will be which leads to the first result:

Proposition 1 *Higher powered incentive schemes reduce the manager's inclination to follow her private preferences in the promotion decision.*

Anticipating M's decision in 2.3 agent A's expected utility is given by

$$U_A = \Pr\left(E\left[a_A \mid s_{A1}\right] - E\left[a_B \mid s_{B1}\right] \geq \frac{\Delta\eta \cdot \Delta w}{\beta\left(k-1\right)}\right) \Delta w - \frac{c}{2} e_{A1}^2.$$

The conditional expectation about agent i's ability is given by

$$E\left[a_i \mid s_{i1}\right] = m_a + \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\varepsilon^2}\left(a_i + e_i + \varepsilon_{i1} - m_a - \widehat{e}_{i1}\right) \tag{2.4}$$

where $\widehat{e}_{i1}$ denotes M's belief about agent i's equilibrium effort choice.[3] Substituting M's conditional expectation into agents' objective function yields

$$EU_A = \Phi\left(\frac{e_{A1} - e_{B1} - \widehat{e}_{A1} + \widehat{e}_{B1} - \frac{\Delta\eta \cdot \Delta w}{\beta(k-1)} \frac{\sigma_a^2 + \sigma_\varepsilon^2}{\sigma_a^2}}{\sqrt{\sigma_a^2 + \sigma_\varepsilon^2}}\right) \Delta w - \frac{c}{2} e_{A1}^2$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. If an internal equilibrium exists[4], the agents' optimal symmetric effort choices are therefore given by

$$e_{A,B}^* = \frac{\Delta w}{c\sqrt{2\pi\left(\sigma_a^2 + \sigma_\varepsilon^2\right)}} \exp\left[-\frac{\Delta\eta^2 \cdot \Delta w^2}{2\beta^2\left(k-1\right)^2} \frac{\left(\sigma_a^2 + \sigma_\varepsilon^2\right)}{\sigma_a^4}\right]. \tag{2.5}$$

From this equation we directly obtain our second result:

Proposition 2 *The agents' equilibrium effort levels are strictly decreasing in the degree of favoritism $|\Delta\eta|$ and strictly increasing in the power of managerial incentives β.*

[3] For the conditional expectation of normally distributed random variables see for instance DeGroot (1970) p. 167.

[4] Existence can be assured when c is sufficiently large as the objective functions are then strictly concave.

Favoritism towards subordinates does not only reduce future profits as on average less able agents are promoted. It also lowers ex-ante profits as the agents anticipate that promotion decisions are not entirely driven by performance considerations. This mechanism weakens the link between performance and rewards and, in turn, makes exerting high efforts less attractive.

To illustrate the relationship between managerial incentives and promotion quality we derive the ex-ante probability of promoting the more able agent. Assume w.l.o.g. that $\Delta\eta > 0$, i.e., agent B is favored by M. The likelihood that indeed the agent is promoted who is expected to be more able is given by

$$1 - \Pr\left(0 < E_{an}\left[a_A \mid s_{A1}\right] - E_{an}\left[a_B \mid s_{B1}\right] < \frac{\Delta\eta \cdot \Delta w}{\beta\left(k-1\right)}\right).$$

Inserting the conditional expectation (2.4) and simplifying yields that in equilibrium this probability is equal to

$$\frac{3}{2} - \Pr\left(\Delta a + \Delta\varepsilon < \frac{\sigma_a^2 + \sigma_\varepsilon^2}{\sigma_a^2} \frac{\Delta\eta \cdot \Delta w}{\beta\left(k-1\right)}\right).$$

As from an ex-ante perspective $\Delta a + \Delta\varepsilon$ is normally distributed with mean 0 and variance $2\sigma_a^2 + 2\sigma_\varepsilon^2$ this probability is

$$\frac{3}{2} - \Phi\left(\frac{1}{2\sigma_a^2} \frac{\Delta\eta \cdot \Delta w}{\beta\left(k-1\right)}\right). \tag{2.6}$$

The function is monotonically increasing in β but becomes flat if β is sufficiently large.[5] Figure 2.1 displays plots of function (2.6) for different degrees of favoritism $\Delta\eta$.[6] Note that even very low powered incentive schemes suffice to generate substantial efficiency gains when $\Delta\eta$ is not too large. Of course, when the impact of favoritism is large, higher values of β become necessary to reduce the bias.

[5] Note that $\lim_{\beta\to\infty} \frac{\partial\left(\frac{3}{2} - \Phi\left(\frac{1}{2\sigma_a^2}\frac{\Delta\eta\cdot\Delta w}{\beta(k-1)}\right)\right)}{\partial\beta} = 0$.

[6] Figure 2.1 shows graphs for $w = 10$, $k = 3$ and $\sigma_a^2 = 5$ and $\Delta\eta = 0.1$, 0.2, or 0.4.

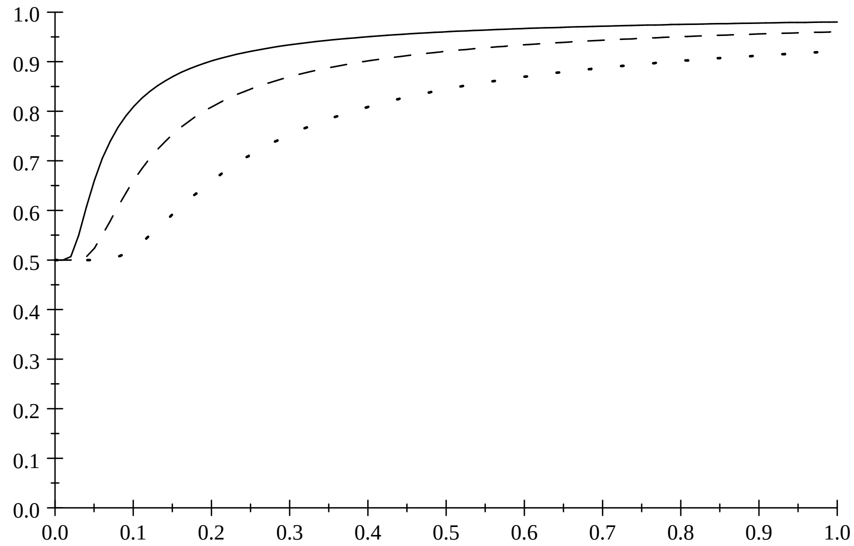

Figure 2.1: Probability that more able agents are promoted as a function of β. Graphs for $\Delta\eta = 0.1$ (solid), $\Delta\eta = 0.2$ (dashed) and $\Delta\eta = 0.4$ (dotted).

2.5 Data and Hypotheses

Our data source is a 2006 employer-employee matched survey conducted by the Great Place to Work Institute and the German Federal Ministry of Labor and Social Affairs. The dataset is a representative sample of 305 German firms employing a minimum of 20 workers. For each firm the management provided company-level information on organizational facts, strategic goals and corporate values as well as on various management practices and the structure of compensation. Most of this information is provided separately for managers and workers in each firm.[7]

In addition to this firm-level information, a representative employee-survey was conducted in each sampled firm yielding over 36,000 observations in total. The employee survey includes 58 standardized items to be answered on a 5-point Likert scale which are designed to measure the level of trust, pride, and cooperation within firms. More precisely the items focus on the

[7] More specifically, answers were provided for employees in supervisory function and the largest group of nonmanagerial employees, i.e. the core occupational group.

relationship among employees, between employees and management, and on the work environment.

Due to the random sampling process the 305 firms are almost evenly spread across the different industries in Germany. The majority of the sampled firms are small or medium sized. While the average number of employees is 430, the median lies at 157. However, roughly 10% of the firms employ more than 1,000 workers including the largest firm in the sample with 14,000 workers.

The management survey includes detailed information on the structure of incentive pay in each firm. Each management representative stated whether wages for managers and workers in the corresponding firm include a performance related pay component. For both, managers and workers, we know the share of the average wage (in %) determined by performance related pay (henceforth PRP).[8]

Figure 2.2 gives a descriptive overview of PRP usage across industries displaying the share of firms using PRP for managers and workers. While only less than half of all sampled firms use variable pay components for workers, the use of manager PRP varies from only 16% in the public sector to 90% in financial services. In total 168 out of 296 (57%) firms use PRP for their managers.[9]

In addition to the information on the strength of performance related pay components, the management survey also includes information on whether there is a gain sharing scheme or managers hold company assets. In contrast to manager PRP, this information is only provided as a binary variable which we label as *Manager Gain Sharing.* 36 out of 295 (12%) firms used such gain sharing plans for their managers. Together with manager PRP this variable will serve as our main independent variable in the upcoming analysis.[10]

[8] To be precise, the items are "*Does the compensation of the employees encompass a performance-based part? (yes/no)*" and "*How big is the variable share on average (in %)*" and "*What are the shares of the following measures of success in this variable compensation component? (company success, success of the organizational unit (team, working group), personal performance, or other)*". It is important to note that this does not refer to the actual payments in the studied year but the general structure of the compensation scheme.

[9] 9 out of the 305 sampled firms did not provide information on PRP.

[10] Note that manager PRP and manager gain sharing are two different ways of tying man-

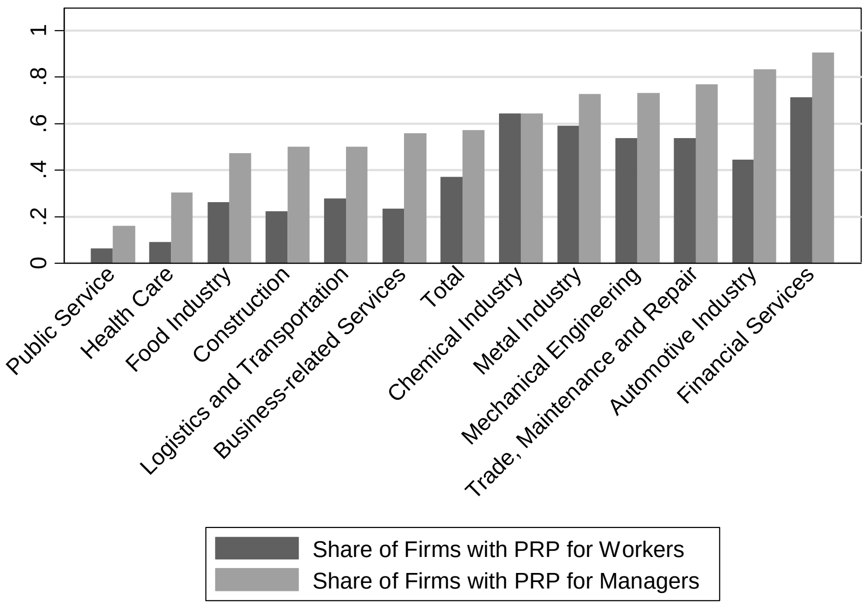

Figure 2.2: Utilization of Performance related Pay across German Industries

Typically, it is very hard to assess the quality of promotion decisions empirically. One reason is that the counterfactual, i.e., the performance of the non-promoted employees on the considered position is never observable. Furthermore, personnel records such as personal assessments or employees' past performances (data that are usually hard to obtain) may not reveal which candidate best meets the requirements for the specific position to be filled. However, employees in a company are in a good position to judge whether indeed the best people are promoted.

Complementing the firm level information provided by the management, we therefore exploit the employee surveys conducted in each firm to measure the quality of promotion decisions.[11] The survey item *"Promotions go to those who best deserve them"* measures the perceived promotion quality

agers' pay to company performance. The two variables show only a weak and insignificant correlation of $r = 0.07$.

[11] In firms with less than 500 employees all employees were asked to participate. In larger firms a representative 500-employee sample was drawn. For sampling details see Hauser et al. (2008).

within a firm. The item is to be answered on a 5-point Likert scale ranging from 1 "almost always untrue" to 5 "almost always true" and refers to the company as a whole.

Figure 2.3 shows the distributions of the proportion of employees per firm who agree with the statement "*Promotions go to those who best deserve them*" (by having chosen a 4 or a 5 (Top Boxes) on the Likert scale). The upper (lower) panel illustrates the distribution for firms without (with) manger PRP.[12] According to the Wilcoxon/Mann-Whitney rank-sum test the data do not stem from the same distribution ($p = 0.000$). The descriptive evidence for firms with and without manager gain sharing plans in figure 2.4 looks similar. Again we can reject the null hypothesis that data come from the same distribution ($p = 0.000$). These patterns of empirical distributions are in line with our theoretical prediction, suggesting that the quality of promotion decisions is indeed higher when firms provide monetary incentives for their managers.

In section 2.6 we test this prediction by running ordered probit regressions. The existence and strength of managers' pay for performance and the existence of gain sharing plans are our main explanatory variables. In section 2.6.2 we address firm heterogeneity by using detailed firm-level information about human resource and management practices and demographic information. Tables 2.5 and 2.6 in the Appendix display descriptive statistics for all variables on the firm and individual level.

2.6 Results

2.6.1 Managerial Incentives and Favoritism

In this section we estimate the relation between managerial incentives and our measure of promotion quality controlling for key firm and employee characteristics. Since employees stem from 305 firms, we cluster the standard errors on firm level. The results of the ordered probit regression are presented in

[12]Since employees' answers are not independent, we aggregate on firm level to report descriptive statistics.

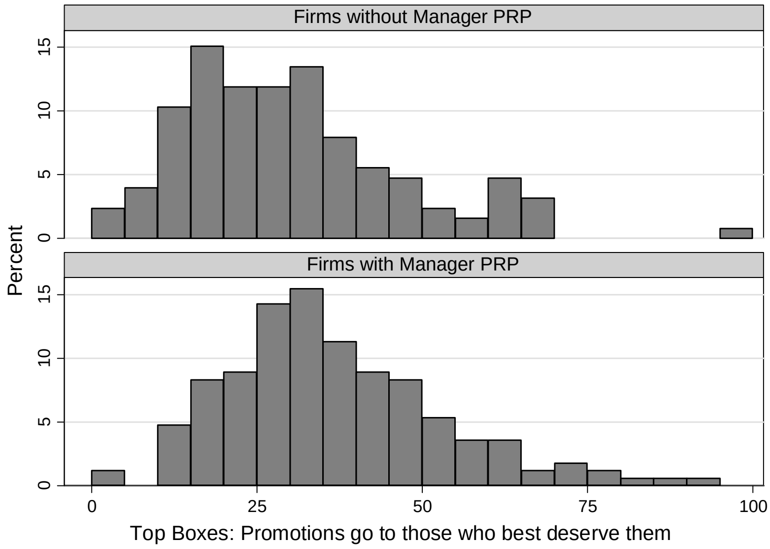

Figure 2.3: Histogram of Proportions of positive Answers in Firms with and without Manager PRP

Figure 2.4: Histogram of Proportions of positive Answers in Firms with and without Manager Gain Sharing Plans

table 2.1. Starting with a basic model we successively add further controls. In column 1 we regress the survey item on a dummy indicating whether a firm provides performance related pay to its managers and a set of standard firm controls including two firm size dummies, 11 industry dummies, and a dummy indicating whether a firm has established a works council.[13] Manager PRP is positively related to the measure of promotion quality, so employees in firms with manager PRP are significantly more likely to state a very strong agreement with the item "*Promotions go to those who best deserve them*" and significantly less likely to state a strong disagreement. In column 2 we alternatively use the presence of a manager gain sharing plan as explanatory variable and again find a highly significant and positive coefficient. Moreover, when including both explaining variables in specification (3), coefficients remain stable in statistical significance, indicating that both are separately related to the quality of promotion decisions, i.e., firms which use both instruments have a higher perceived quality of promotions as compared to firms which use only one of them.[14]

[13]According to German law, firms are obliged to set up a works council when this is demanded by employees.

[14]There is no detectable interaction effect between manager PRP and manager gain sharing when an interaction term is included.

Dependent Variable	"Promotions go to those who best deserve them"				
	Ordered Probit				Probit
	(1)	(2)	(3)	(4)	(5)
Manager PRP	0.254***		0.250***	0.241***	0.0828***
	(0.0560)		(0.0571)	(0.0593)	(.0218)
Manager Gain S.		0.217**	0.209**	0.241***	0.0972 ***
		(0.0958)	(0.0941)	(0.0828)	(.0292)
Internal Staffing (%)				0.00105	0.000485*
				(0.000650)	(.000256)
Works Council	-0.163***	-0.165***	-0.166***	-0.145***	-0.0516***
	(0.0509)	(0.0526)	(0.0488)	(0.0460)	(.0184)
Socioeconomics	No	No	No	Yes	Yes
Observations	28,373	28,373	28,373	25,917	25,917
Clusters (Firms)	294	294	294	274	274
Chi^2	517.2	431.2	550.9	924.6	771.6
Pseudo R^2	0.02	0.02	0.02	0.03	0.05

*** $p<0.01$, ** $p<0.05$, * $p<0.1$, clustered standard errors on firm level in parentheses

Column 1-4: Ordered Probit regression, column 5: Probit regression, marginal effects

Further controls: 2 firm size dummies and 11 industry dummies

Socioeconomics include tenure, sex, occupational status, job type (blue collar, white collar)

Reference category: full-time employee in a small firm (0-99 employees) in the food industry

Table 2.1: The Relationship between Managerial Incentives and Favoritism

Probability of Promotion Quality (PQ)	Pr(PQ=1)	Pr(PQ=2)	Pr(PQ=3)	Pr(PQ=4)	Pr(PQ=5)
Marginal Effects for Column 1					
Dummy Manager PRP	-0.0520	-0.0379	0.00276	0.0607	0.0265
Dummy Manager Gain S.					
Internal Staffing (%)					
Works Council	0.0299	0.0251	0.00368	-0.0390	-0.0196
Marginal Effects for Column 2					
Dummy Manager PRP					
Dummy Manager Gain S.	-0.0382	-0.0335	-0.00756	0.0515	0.0277
Internal Staffing (%)					
Works Council	0.0303	0.0253	0.00377	-0.0394	-0.0200
Marginal Effects for Column 3					
Dummy Manager PRP	-0.0511	-0.0374	0.00265	0.0599	0.0260
Dummy Manager Gain S.	-0.0367	-0.0324	-0.00707	0.0498	0.0263
Internal Staffing (%)					
Works Council	0.0303	0.0256	0.00381	-0.0397	-0.0199
Marginal Effects for Column 4					
Dummy Manager PRP	-0.0472	-0.0371	0.00160	0.0585	0.0242
Dummy Manager Gain S.	-0.0397	-0.0381	-0.0105	0.0582	0.0301
Internal Staffing (%)	-0.000196	-.0000165	-8.053e-06	0.000257	0.000112
Works Council	0.0255	0.0229	0.00361	-0.0353	-0.0167
Ordered probit marginal effects column 1-4, table 2.1					

Table 2.2: Ordered Probit Marginal Effects

In specification (4) we control for the socioeconomic background (tenure, age, and education) and the percentage of managerial positions filled internally (*Internal Staffing*), since higher internal staffing quotas may be correlated with our explaining variables and cause an upward bias in the respective answers. While we see that manager gain sharing plans as well as higher internal staffing quotas are also significantly positively related to higher promotion quality, the coefficient of our main variables of interest remains stable and highly significant. Table 2.2 shows the ordered probit marginal effects of the estimates from column 1-4 in table 2.1. The bottom panel in table 2.2 shows that employees in firms which apply manager PRP, are 2.4% (5.9%) more likely to choose 5 (4) on the Likert scale and are less likely to make negative statements (choose 1 or 2). Similar, employees in firms with managerial gain sharing plans are 3.0% (5.8%) more likely to make the respective statement.

To develop a more intuitive way for evaluating the economic significance of the relationship in question, we dichotomized the dependent variable and created a dummy variable which has value 1 if an employee chooses one of the two highest levels of agreement to the statement that promotions go to those who best deserve them. We regress this dummy on the explanatory variables from the specification in column 4 in table 2.1 in a simple binomial probit regression model. Column 5 in table 2.1 shows the marginal effects of the binomial probit estimation.[15] A worker is 8.3% more likely to agree that promotions are based on merit if her firm uses manger PRP and a further 9.7% more likely when manager gain sharing is used.

But what does this increase in probability mean for a firm? To obtain an additional economic interpretation we aggregate the data and run an OLS regression on firm level with explanatory variables from table 2.1, column 1-4. The percentage of employees who agree that promotions are based on merit (i.e., ticking 4 or 5) serves as dependent variable. The results are shown in the Appendix in table 2.7. In firms with manager PRP (manager gain sharing plan) 5.7% (6.7%) more employees have a positive opinion about their firm's

[15] Running linear models on the discrete outcome variable, as considered by Angrist and Pischke (2009), yields similar results.

promotion practice.[16] The predicted fraction of employees agreeing to that statement at the mean of all other explanatory variables is equal to 30.7% (33.3%) when there is no manager PRP (gain sharing). This rate increases to 36.4% (40%) when manager PRP (gain sharing) is applied. Overall, in firms which apply one or both instruments, substantially more employees believe that promotions are based on merit.

In the next step we consider the effect of the strength of managerial incentives. The key independent variable is now the average percentage of manager PRP. The theoretical model in section 2.3 predicts that higher managerial incentives should lead to less favoritism. However, as indicated by figure 2.1, for higher manager PRP, additional gains in promotion quality should decline in the size of manager PRP. For instance, when preferences for favoritism are not too strong, already weak managerial incentives lead to good promotion decisions. Any further increase in managerial incentives may provide only little further improvement.

In table 2.3 we therefore regress our main dependent variable on the strength of manager PRP (column 1 & 2). In column 3 we include a squared term, while column 4 includes 3 interval dummies for the strength of manager PRP with the reference category being firms without manager PRP.

The coefficients in columns 1 & 2 in table 2.3 again indicate a positive relationship between the two incentives schemes and the likelihood of a positive statement. The negative and significant square term in column 3 and 4 indeed reveals decreasing returns to performance pay. In column 5 we use dummies for firms which use weak (1%-10%), moderate (11%-20%) or strong (>20%) manager PRP, where firms without managerial PRP serve as base category. The three dummy coefficients are all statistically significant and similar in size. Promotion quality in firms with moderate or high managerial incentives is higher than in firms from the base category but as high as in firms with low incentives.[17] In sum, the results in table 2.3 fit the theoretical prediction that even weak incentives are associated with substantially higher

[16] Ticking 4 or 5 on the Likert scale we describe as *positive*, ticking 1 or 2 we describe as *negative* judgement.

[17] Note that the coefficients of the interval dummies are not significantly different from one another.

Dependent Variable	"Promotions go to those who best deserve them"				
	(1)	(2)	(3)	(4)	(5)
Manager PRP (%)	0.0045**		0.0119***	0.0109**	
	(0.0022)		(0.0045)	(0.0044)	
Manager PRP^2 (%)			-0.00012**	-0.00011*	
			(6.11e-05)	(6.08e-05)	
Manager PRP 1-10%					0.256***
					(0.0649)
Manager PRP 11-20%					0.240***
					(0.0700)
Manager PRP > 20%					0.199**
					(0.0914)
Manager Gain S.		0.217**	0.312**	0.281***	0.231***
		(0.0958)	(0.0911)	(0.0917)	(0.0770)
Internal Staffing (%)				0.0010	0.0011
				(0.0006)	(0.0006)
Works Council	-0.147***	-0.165***	-0.138***	-0.130***	-0.150***
	(0.0554)	(0.0526)	(0.0518)	(0.0503)	(0.0462)
Socioeconomics	Yes	Yes	Yes	Yes	Yes
Observations	27,948	28,373	27,948	25,840	25,840
Clusters (Firms)	288	294	288	271	271
Chi^2	458.2	431.2	499.0	851.6	922.9
Pseudo R^2	0.02	0.02	0.02	0.03	0.03

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, clustered standard errors on firm level in parentheses
Ordered Probit regression, further controls: 2 firm size dummies and 11 industry dummies
Reference category: 0-99 employee firm in the food industry
Reference category in model (4): Firms without manager PRP
Socioeconomics include tenure, sex, occupational status, job type (blue collar, white collar)

Table 2.3: The Relationship between Managerial Incentives and Favoritism - Quadratic and Nonparametric Models

promotion quality.

2.6.2 Robustness and Potentially Omitted Variables

Of course it is important to discuss to what extent the observations reported in the above may be due to unobserved firm heterogeneity rather than a causal relationship between incentive schemes and promotion quality. A competing non-causal explanation for our cross sectional evidence might be that the management's willingness to use modern Human Resource Management practices causes better promotion quality, which in turn is correlated with our explanatory variables. For example, a firm may have invested more in setting up a professional HR department which at the same time advocates the use of PRP and performance appraisal procedures that help to promote the most able employees. Or, a firm with higher growth rates may have a higher willingness to share profits with managers and at the same time provides more promotion opportunities which leads the workforce to conclude that many capable employees can be promoted. While our dataset does not provide a variable that is suitable in a sensible manner as an instrument in an IV regression (all possible candidates violate the exclusion restriction without imposing untenable assumptions), we have rich information on other HR practices used by the firms. By controlling for management and HR practices we can provide strong evidence that rules out many potential non-causal interpretations of the results described in the above. The key advantage of our data is that we have two combined data sources and, in a narrow sense, the management practices are exogenous when viewed from the perspective of an individual employee whose perceptions we measure and use as dependent variable.

We extend the specification from column 4 in table 2.1 by stepwise adding further controls in table 2.4. In column 1 we add a dummy for the presence of performance related pay for non-managerial employees and expect a positive sign, since worker PRP needs a system to evaluate performance and this may induce higher promotion quality. However, worker PRP is not correlated with perceived promotion quality. Moreover, the coefficients for manager

Dependent Variable	"Promotions go to those who best deserve them"				
	(1)	(2)	(3)	(4)	(5)
Manager PRP	0.212***	0.188**	0.187**	0.149**	0.161**
	(0.0739)	(0.0784)	(0.0762)	(0.0658)	(0.0651)
Manager Gain S.	0.227***	0.235***	0.235***	0.275***	0.265***
	(0.0769)	(0.0770)	(0.0770)	(0.0727)	(0.0770)
Worker PRP	0.0574	0.0669	0.0678	0.0723	0.0416
	(0.0618)	(0.0626)	(0.0618)	(0.0581)	(0.0594)
Upsize		0.122**	0.122**	0.115**	0.0894*
		(0.0566)	(0.0562)	(0.0521)	(0.0505)
Downsize		0.00743	0.00799	-0.00236	0.0361
		(0.0566)	(0.0567)	(0.0611)	(0.0561)
Female Career			-0.00902	-0.0389	-0.0460
			(0.0712)	(0.0737)	(0.0709)
Job Rotation				0.108**	0.135***
				(0.0495)	(0.0498)
High Wage					0.0339
					(0.0461)
Low Wage					-0.179
					(0.115)
Observations	25,917	25,917	25,917	23,044	22,279
Clusters (Firms)	274	274	274	252	235
Chi^2	933.3	975.6	1003.9	1002.7	1322.8
Pseudo R^2	0.03	0.03	0.03	0.03	0.03

*** $p<0.01$, ** $p<0.05$, * $p<0.1$, clustered standard errors in parentheses
Ordered Probit regression, further controls: 2 firm size dummies and 11 industry dummies. Socioeconomics include tenure, sex, occupational status, and job type (blue collar, white collar). Reference category: full-time employee in a small firm (0-99 employees) in the food industry

Table 2.4: The Relationship between Managerial Incentives and Favoritism - Controls I

PRP and manager gain sharing are basically unaffected. If our results were largely driven by the general willingness to use modern HR practices we should observe a substantial drop in the coefficients which is not the case.

In the next step we add information on whether the firm up- or downsized within the last 3 years. In particular, both promotion opportunities and the willingness to share profits may increase in firms with high growth rates. The dummy *Upsize* (*Downsize*) equals 1 if the number of employees increased (decreased) by more than 5% in the last 3 years. As can be seen in column (2) *Upsize* is positively correlated with the promotion quality, though this effect vanishes in the following richer specifications. Again the coefficients for manager PRP and gain sharing are hardly affected. We additionally control for the existence of a specific woman career plan and job rotation program, which both potentially affect promotion decisions and promotion quality. Ortega (2001), Arya and Mittendorf (2004) and Eriksson and Ortega (2006) argue that firms learn about employees' productivity and specific capabilities via job rotation and that employer learning increases promotion quality. The effect of woman career plans is ambiguous. The selective promotion of high ability women might increase the promotion quality, but male employees may feel discriminated and thus perceive a lower promotion quality. In column (3) and (4) we add dummies for both measures. Only job rotation is positively correlated with our measure of promotion quality. In column (5) we add dummies for companies paying above or below the respective union wages, where companies with above union wages show higher perceived promotion quality.

We add further controls in specifications reported in table 2.8 in the Appendix. For instance, we control for the average number of days of further employee training as this should be a suitable proxy for a firm's willingness to invest in human resources (which may affect promotion quality and the willingness to share profits). As several firms did not report this key figure the number of observations drops substantially. However, it is interesting to note that the coefficients for manager PRP and gain sharing increase substantially in size and remain highly significant. Finally, we control for the general work satisfaction of employees as higher perceived promotion quality may be simply due to higher work satisfaction. As can be seen in column (8), work satisfaction is indeed highly correlated with promotion quality, but does not affect the association between manager PRP or gain sharing and perceived promotion quality.

2.7 Conclusion

When performance is not perfectly observable promotion decisions are frequently based on subjective performance evaluations. If managers have personal preferences for certain workers they have an incentive to distort performance ratings which promotes favored workers rather than most able workers. We theoretically show that favoritism can reduce company profits not only by putting the wrong people into management positions but also by reducing incentives for workers to exert effort in promotion tournaments. Managerial incentives can constrain favoritism in promotion decisions by realigning managers' and firm's interest and even rather weak managerial incentives may generate strong efficiency gains.

We empirically tested the theoretical prediction using an unique representative matched employer-employee dataset. The empirical analysis confirms our theoretical result as promotion quality is significantly higher in firms in which managers receive performance related pay or participate in gain sharing plans. Indeed this higher perceived quality of promotion decisions should translate in higher company performance as, not only the probability that the best and not the best liked employees are promoted increases, but also because stronger merit-based promotions should lead to a more motivated workforce.

2.8 Appendix to Chapter 2

Explanatory Variables: Firm Level

Name	Description	Min	Max	Mean	Sd	N
Manager PRP	=1 if firm uses managerial incentives	0	1	0.57	.50	294
Manager PRP	% of total wage depending on variable payment	0	80	4.56	8.81	292
Manager Gain S.	=1 if managers receive gain sharing/ hold firm assets	0	1	0.12	0.33	294
Worker PRP	=1 if firm uses worker PRP	0	1	0.37	.48	294
Works Council	=1 if works council in the firm	0	1	0.60	0.49	294
Internal Staffing	% of managerial vacancies filled with internal hires	0	100	52.20	37.78	274
Industries	12 industry dummies	1	12	-	-	305
Firm Size	3 firm size dummies: 0-99, 100-499 and >500 employees	1	3	-	-	305
Internal Promotion	% of positions filled with internal staff	0	100	52.19	37.8	274
Upsize	dummy for a >5% employee upsize from 2003-2006	0	1	0.33	0.47	305
Downsize	dummy for a >5% employee downsize 2003-2006	0	1	0.28	0.45	305
Female career	dummy for female career planing	0	1	0.065	0.25	294
Job Rotation	dummy for a firm wide job rotation program	0	1	0.48	0.50	275
High Wage	dummy for wage level above corresponding union wage	0	1	0.33	0.47	268
Low Wage	dummy for wage level below corresponding union wage	0	1	0.037	0.19	268
Further training	average number of training days per employee and year	0	321	14.10	44.25	171
Recruiting	dummy for strategically recruiting experts	0	1	0.26	0.44	275

Table 2.5: Descriptive Statistics of Variables on Firm Level

Explanatory Variables: Individual Level

Name	Description	Min	Max	Mean	Sd	N
Tenure	employee's firm tenure	1	>21	11.2	7.11	34,697
Male	gender dummy	0	1	0.57	0.50	34,697
Leader	dummy for managerial employees	0	1	0.28	0.45	34,592
Part-time	dummy for part-time occupation	0	1	0.17	0.38	34,731
Educ	categorical variable for 9 education types	0	8	-	-	33,577
Work Satisfaction	general work satisfaction; 7-point Likert scale	1	7	5.01	1.15	32,848

Table 2.6: Descriptive Statistics of Variables on Individual Level

Dependent Variable	"Promotions go to those who best deserve them" Proportion of employees stating agreement (0-100)			
	(1)	(2)	(3)	(4)
Manager PRP	7.110***		6.801***	5.665***
	(2.294)		(2.275)	(2.088)
Manager Gain Sharing		8.093***	7.572***	6.787***
		(2.627)	(2.593)	(2.593)
Internal Staffing (%)				0.0533**
				(0.0249)
Works Council	-6.775***	-6.150***	-6.521***	-4.433**
	(2.044)	(2.145)	(2.048)	(2.097)
Socioeconomics	No	No	No	Yes
Constant	37.63***	38.93***	36.44***	51.29***
	(4.274)	(4.285)	(4.360)	(5.609)
Observations	294	294	294	274
R^2	0.19	0.18	0.21	0.31

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, robust standard errors in parentheses

OLS regression: further controls: 2 firm size dummies and 11 industry dummies

Reference category: 0-99 employee, firm in the food industry

Socioeconomics include average tenure & share of males among respondents

Table 2.7: The Relationship between Managerial Incentives and Favoritism - OLS Analysis

Dependent Variable	"Promotions go to those who best deserve them"		
	(1)	(2)	(3)
Manager PRP	0.165**	0.320***	0.315***
	(0.0651)	(0.0860)	(0.0818)
Manager Gain Sharing	0.264***	0.378***	0.362***
	(0.0772)	(0.0887)	(0.0960)
Upsize	0.0905*	0.0904*	0.0592
	(0.0505)	(0.0532)	(0.0535)
Downsize	0.0347	-0.00614	-0.0224
	(0.0554)	(0.0734)	(0.0663)
Job Rotation	0.134***	0.149***	0.139**
	(0.0497)	(0.0569)	(0.0556)
High Wage	0.0345	0.0398	0.0206
	(0.0458)	(0.0580)	(0.0565)
Low Wage	-0.182	-0.784***	-0.631***
	(0.117)	(0.0944)	(0.0960)
Recruiting	-0.0149	-0.0314	-0.0124
	(0.0436)	(0.0484)	(0.0481)
Further Training		0.000526	0.000288
		(0.000453)	(0.000432)
Work Satisfaction			0.306***
			(0.0122)
Observations	22,279	13,571	12,724
Chi^2	1301.6	1882.7	2397.9
Clusters (Firms)	235	139	139
Pseudo R^2	0.03	0.04	0.07

*** p<0.01, ** p<0.05, * p<0.1, clustered standard errors on firm level in parentheses
Ordered Probit regression, Further controls: 2 firm size dummies,
11 industry dummies, works council, worker PRP, female career
Socioeconomics include tenure, sex, occupational status, education

Table 2.8: The Relationship between Managerial Incentives and Favoritism - Controls II

Chapter 3

When Higher Prizes Lead to Lower Efforts - The Impact of Favoritism in Tournaments[1]

3.1 Introduction

Since the seminal contribution of Lazear and Rosen (1981), numerous papers have explored the incentive effects of tournaments. One of the most prominent results in the literature is that higher tournament prizes lead to higher efforts. A key assumption in most of these tournament models is that the agent with the highest output always wins the tournament. However, in reality, tournament outcomes are often based on subjective decisions by individuals. For instance, in organizations managers decide upon promotions. Or in sports contests, referees or judges either directly determine the winner or make decisions which crucially affect the tournament outcome. In these settings not only output, but also personal preferences towards the agents may affect the choice of the winner. We show in a simple extension of the standard Lazear/Rosen framework that the existence of favoritism can reverse the relationship between the tournament prize and effort choices. The effect of higher prizes are then twofold: On the one hand, higher prizes make

[1] This chapter is based upon Herbertz and Sliwka (2011).

it still more attractive to win. But on the other hand there are higher incentives for a biased evaluator to pick her favorite agent and as the tournament becomes more uneven incentives are reduced. We show that the latter effect always dominates the former when prizes are beyond a certain threshold, such that efforts then are strictly decreasing in the prize spread.

Prendergast and Topel (1996) find that subjective evaluation distortion induced by favoritism leads to worse job allocation and lower optimal incentives for employees. Higher incentives for the agents increase the supervisor's inclination to distort evaluations. Prendergast (2002) extends this approach by showing that favoritism in subjective performance evaluations can lead to a reverse relationship of risk and incentives. In a risky environment, a supervisor's evaluation is noisy and has little weight since it bears fewer information. Hence, favoritism becomes less important and optimal incentives are higher than in less risky environments. Berger et al. (2011) show, that favoritism in Lazear/Rosen job promotion tournaments leads to lower efforts and lower promotion quality. This effect can be mitigated by proper managerial incentives. We use a similar model as Berger et al. (2011) to derive conditions under which higher tournament prizes lead to lower efforts.

3.2 The Model

Consider a model with a supervisor S and two agents $i = A, B$ who compete in a tournament. The agents choose an unobservable effort level e_i at costs $c(e_i)$ and produce outputs

$$s_i = a_i + e_i + \varepsilon_i$$

where $a_i \sim N(m_a, \sigma_a^2)$ denotes agent i's unknown ability. The error term is also normally distributed with $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. We assume a_i and ε_i to be independent and their distributions to be common knowledge. All players are risk neutral. After the agents have exerted their efforts the supervisor S observes the unverifiable performance signals s_i and picks the winner of the tournament $\phi \in \{A, B\}$. The winning agent receives a tournament prize P.

The supervisor benefits when the more able agent wins the tournament. For instance, a manager earns more when picking the more able candidate for a promotion as profits are higher, or the reputation of a judge in a sport contest is affected by the future performance of the winner. But the supervisor also cares for the well-being of the agents and does so to a different extent for the two agents. Similar to Prendergast and Topel (1996) or Prendergast (2002) we assume that these preferences are measured by a parameter η_i indicating how much the supervisor cares for an agent i. The supervisor's overall utility is thus a function of the chosen winner ϕ and is given by

$$V(\phi) = k \cdot a_\phi + \eta_\phi \cdot P.$$

Hence, k measures the degree of alignment, i.e., the higher k the higher are the incentives for the supervisor to indeed pick the agent she believes to be more able. But when the preference parameters η_i differ strongly she may favor the agent whom she likes more even when this comes along with a lower expected ability of the winner.

3.3 Equilibrium Analysis

We now determine the Perfect Bayesian Equilibrium of the game described above. The supervisor will declare that agent A is the winner when

$$\begin{aligned} E\left[V(A)\mid s_A\right] > E\left[V(B)\mid s_B\right] &\Leftrightarrow \\ E\left[a_A\mid s_A\right] - E\left[a_B\mid s_B\right] &> \frac{\Delta\eta_B \cdot P}{k} \end{aligned} \tag{3.1}$$

where $\Delta\eta_B = \eta_B - \eta_A$. Hence, if agent B is favored (i.e., $\Delta\eta_B > 0$) then A wins the tournament only if she is believed to be sufficiently more able than agent B. In the absence of favoritism ($\Delta\eta_B = 0$) the right hand side in equation 3.1 is zero. In this case, S's decision is solely driven by her expectations about agents' abilities and the agent who is expected to be more able is chosen. Hence, the model boils down to a standard Lazear and Rosen (1981) type tournament. If, however, favoritism matters, S gains additional utility from picking the favored agent. The more S favors an agent, the more likely it is that her decision will not coincide with the ability maximizing decision. The higher k the smaller is this distortion.

Now we can analyze how the supervisor's decision depends on actual performance outcomes. The conditional expectation on A's ability is given by

$$E\left[a_A|\, s_A\right] = m_a + \frac{\sigma_a^2}{\sigma_a^2+\sigma_\varepsilon^2}\left(a_A + e_A + \varepsilon_A - m_a - \hat{e}_A\right) \tag{3.2}$$

where $\hat{e}_{A1}$ denotes S's belief about agent A's equilibrium effort choice.[2] Hence, agent A will be promoted if

$$m_a + \frac{\sigma_a^2}{\sigma_a^2+\sigma_\varepsilon^2}\left(s_A - m_a - \hat{e}_A\right) - m_a - \frac{\sigma_a^2}{\sigma_a^2+\sigma_\varepsilon^2}\left(s_B - m_a - \hat{e}_B\right) > \frac{\Delta\eta_B \cdot P}{k}.$$

This directly leads to the following result:

Lemma 1 *The supervisor picks agent A as the winner of the tournament if and only if*

$$s_A - s_B > \frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B \cdot P}{k} + \hat{e}_A - \hat{e}_B.$$

Even at identical effort levels, A wins the tournament only if she outperforms B with a sufficiently large margin when B is favored by the supervisor. Both competitors take the distortion into account when they choose their efforts. Anticipating S's decision agent A's expected utility is given by

$$\begin{aligned}
&\Pr\left(a_A + e_A + \varepsilon_A - a_B - \hat{e}_B - \varepsilon_B > \frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B \cdot P}{k} + \hat{e}_A - \hat{e}_B\right) P - c\left(e_A\right) \\
&\quad = \Pr\left(e_A - \hat{e}_A - \frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B \cdot P}{k} > a_B - a_A + \varepsilon_B - \varepsilon_A\right) P - c\left(e_A\right).
\end{aligned}$$

As $a_B - a_A + \varepsilon_B - \varepsilon_A$ is a normally distributed random variable with mean

[2] For the conditional expectation of normally distributed random variables see for instance DeGroot (1970) p. 167.

0 and variance $2\left(\sigma_a^2+\sigma_\varepsilon^2\right)$ this is equal to

$$\Phi\left(\frac{e_A-\hat{e}_A-\frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B\cdot P}{k}}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)P-c\left(e_A\right)$$

where $\Phi\left(\cdot\right)$ is the c.d.f. of the standard normal distribution. Suppose for the moment that an internal pure strategy equilibrium exists. If this is the case, the agents' optimal effort choices are characterized by the first order condition

$$\phi\left(\frac{e_A-\hat{e}_A-\frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta B\cdot P}{k}}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)\frac{1}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}P-c'\left(e_A\right)=0. \tag{3.3}$$

In equilibrium $e_A=\hat{e}_A$ and as the left hand side is equal for A and B we can show the following result:

Proposition 1 *If* $\frac{P}{\sqrt{8\pi\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\exp\left(-\frac{1}{2}\right)<\inf_e c''\left(e\right)$ *there is a unique pure strategy equilibrium in which both agents exert effort level*

$$e^*=c^{-1}\left(\phi\left(-\sqrt{\frac{1}{2}\frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^4}}\frac{\Delta\eta_B P}{k}\right)\frac{P}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right) \tag{3.4}$$

and the supervisor declares that agent A is the winner if $s_A-s_B>\frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B\cdot P}{k}$.

Proof: See the appendix.

Thus, as in a standard Lazear/Rosen type tournament a symmetric equilibrium exists if the cost functions are sufficiently convex.[3] It is straightforward to see that in this equilibrium, the efforts are decreasing in the degree of favoritism $\Delta\eta_B$ as $\phi\left(x\right)$ is symmetric and single peaked at 0. This is the well known result that efforts are lower in asymmetric tournaments. However, the effect of the tournament prize is less evident as there are two countervailing effects: On the one hand winning is more attractive for higher values of P, but on the other hand, the tournament becomes less fair the higher P and

[3] Compare Lazear and Rosen (1981), p. 845, or Bhattacharya and Guasch (1988), p. 871.

this has a negative effect on incentives. Taking the first derivative of the equilibrium effort (3.4) we obtain that $\frac{\partial e}{\partial P} > 0$ is equivalent to

$$-\phi'\left(-\sqrt{\frac{\sigma_a^2+\sigma_\varepsilon^2}{2\sigma_a^4}}\frac{\Delta\eta_B P}{k}\right)\sqrt{\frac{\sigma_a^2+\sigma_\varepsilon^2}{2\sigma_a^4}}\frac{\Delta\eta_B P}{k}+\phi\left(-\sqrt{\frac{\sigma_a^2+\sigma_\varepsilon^2}{2\sigma_a^4}}\frac{\Delta\eta_B P}{k}\right)>0.$$

Using that for normally distributed random variables $\phi'(x) = -x\phi(x)$ this simplifies to $P^2 < 2\frac{\sigma_a^4}{\sigma_a^2+\sigma_\varepsilon^2}\frac{k^2}{\Delta\eta_B^2}$ and we obtain our key result:

Proposition 2 *Too high tournament prizes lead to a reduction of the agents' efforts: there is a threshold value* $\bar{P} = \sqrt{\frac{2\sigma_a^4}{\sigma_a^2+\sigma_\varepsilon^2}}\frac{k}{\sqrt{\Delta\eta_B^2}}$ *for* P *such that the equilibrium effort is strictly decreasing in* P *for* $P > \bar{P}$*. The threshold* $\bar{P}$ *is strictly increasing in the degree of alignment* k *and strictly decreasing in the degree of favoritism* $\Delta\eta_B$*.*

If the tournament prize is high, much is at stake for the agents and the supervisor. When the supervisor favors one of them, the temptation becomes larger to distort the choice of a winner if the prize is high. The supervisor will still pick the agent with the higher performance, if this agent has outperformed the favored competitor to a sufficiently high degree. But this margin is increasing in the tournament prize. This in turn reduces the effort incentives for both the advantaged and the disadvantaged agent. For sufficiently high tournament prizes this distortion effect always outweighs the direct incentive effect as both agents will anticipate that the non-favored agent will have little chances to win.

If σ_ε^2 is high, the signal becomes less valuable in estimating agents ability and favoritism becomes more important in the supervisors promotion decision. Hence, $\bar{P}$ decreases in σ_ε^2. This finding differs from Prendergast (2002) who finds that a risky environment reduces distortions by favoritism in subjective evaluations. In this setting, signal noise reduces the supervisor's weight in performance evaluation and hence limits the potential favoritism by the supervisor. In our model the signal noise σ_ε^2 does not reduce the importance of the supervisor's evaluation, but the importance of an observed difference in signals for the supervisor's promotion decision. If the signals bear less information about the agents' abilities, personal preferences become more important in the promotion decision.

On the other hand if ability noise σ_a^2 is high, signals carry more information, since their fluctuations are more likely due to ability differences than to noise. If signals become more important for promotions, agents have stronger incentives to provide effort for increasing their performance signal. Hence $\bar{P}$ increases in σ_a^2.

3.4 Conclusion

We show that favoritism in tournaments can lead to a reversed relationship between the tournament prize and equilibrium efforts. The higher the prize, the less likely it is that a supervisor will promote the more able agent, because more is at stake for her personal favorite. Large stakes are common in sports, but also in promotion tournaments for well paid jobs, such as executive positions. Especially in the latter, large prizes are accompanied by subjective evaluations and social ties which are crucial for CEO appointments. If it is not possible to eliminate subjectivity in tournaments, tournament prizes should not be too high for avoiding demotivation of competing agents.

3.5 Appendix to Chapter 3

Proof of Proposition 1:

A sufficient condition for existence is that objective functions are strictly concave. This is the case if the second derivative of the objective function is negative for any effort level e_A

$$\phi'\left(\frac{e_A-\hat{e}_A-\frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B\cdot P}{k}}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)\frac{1}{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}P<c''\left(e_A\right)\quad\forall e_A.$$

Using that $\phi'(x)=-x\phi(x)$

$$-\phi\left(\frac{e_A-\hat{e}_A-\frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B\cdot P}{k}}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)\frac{e_A-\hat{e}_A-\frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B\cdot P}{k}}{\left(2\left(\sigma_a^2+\sigma_\varepsilon^2\right)\right)^{1.5}}P<c''\left(e_A\right)\tag{3.5}$$

we show that the lhs is bounded from above. To see that let $y=e_A-\hat{e}_A-\frac{\sigma_a^2+\sigma_\varepsilon^2}{\sigma_a^2}\frac{\Delta\eta_B\cdot P}{k}$. The lhs becomes

$$-\phi\left(\frac{y}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)\frac{y}{\left(2\left(\sigma_a^2+\sigma_\varepsilon^2\right)\right)^{1.5}}P.\tag{3.6}$$

This function has two extreme points as its first order condition is given by

$$-\phi'\left(\frac{y}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)\frac{y}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}-\phi\left(\frac{y}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)=0$$

and simplifies to

$$\Leftrightarrow\phi\left(\frac{y}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)\frac{y^2}{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}-\phi\left(\frac{y}{\sqrt{2\left(\sigma_a^2+\sigma_\varepsilon^2\right)}}\right)=0$$

and hence to $y^2=2\left(\sigma_a^2+\sigma_\varepsilon^2\right)$. The lhs in equation 3.6 is maximized at

$y = -\sqrt{2\left(\sigma_a^2 + \sigma_\varepsilon^2\right)}$. We obtain the maximum of

$$\frac{P}{\sqrt{8\pi}\left(\sigma_a^2 + \sigma_\varepsilon^2\right)} \exp\left(-\frac{1}{2}\right).$$

Therefore, a pure strategy equilibrium always exists if

$$\frac{P}{\sqrt{8\pi}\left(\sigma_a^2 + \sigma_\varepsilon^2\right)} \exp\left(-\frac{1}{2}\right) < \inf_e c''\left(e_A\right).$$

■

Chapter 4

Social Ties, Incentives, and Adverse Selection[1]

4.1 Introduction

Employees in organizations are involved or confronted with social ties between colleagues, supervisors or fellow employees. For several decades, researchers have been analyzing social networks predominantly finding positive effects, such as increased inner firm information flow, trust and productivity, as well as reduced search costs in the labor market (see Granovetter (2005) for a sociological and Rauch and Casella (2001) for an economical research overview). However, strong relationships created by family ties, professional long-term relationships, or fraternity membership, may also foster nepotistic behavior, which creates disadvantages to the unrelated employees, the so-called outsiders, and inefficient decisions in firms.

In this paper, we experimentally investigate whether high ability outsiders adversely select away from strong personal relationships and whether selection can be mitigated by efficiency enhancing managerial incentives which should weaken nepotistic behavior.

In the first stage of the experiment, an unrelated agent can choose between a riskless outside option and entering an organization with two unknown

[1] This chapter is based upon Herbertz (2011).

players who are friends. This resembles the situation that applicants face if they decide for or against working for an organization. If the outside agent chooses to enter the organization, one of the two unknown players distributes a pie among her friend and the outsider. Favoring her friend yields an inefficient allocation where the outside agent receives a slightly lower payoff compared to her outside option. Favoring the outsider yields an efficient allocation where the outsider receives considerably more compared to her outside option.

The results illustrate, that in the absence of managerial incentives, outsiders fear nepotistic behavior by the manager and tend to choose the riskless outside option. If the manager receives a variable payment, where her compensation is tied to the sum of the total payoffs, nepotism is costly. In this case, the outside agent is more inclined to enter the organization, since she hopes for an efficient pie distribution. As it turns out, agents are right in doing so, since managerial incentives indeed reduce managers' favorable behavior towards their friends.

The experimental results suggest that strong personal relationships in organizations might come at the cost of the adverse selection of high ability workers. Transparent managerial incentives help to avoid the adverse selection of high ability employees due to the crowding out of nepotistic behavior.

The paradigmatic case for organizations with strong social ties are family firms which are very prominent in western stock markets. In most studies, the term *family firm* denotes firms where members of one family own substantial firm shares and consequently play a major role in the board of directors or even the board of management. 35% of S&P 500 firms are dominated by families, where 15% of these firms are controlled by founders and 30% are controlled by the founders' descendents CEOs (Anderson and Reeb (2003)).[2] One main research finding is the underperformance of firms headed by descendent CEOs, compared to firms with professional or founder CEOs (Bennedsen et al. (2007), Pérez-González (2006), Vilallonga and Amit (2006)

[2]Figures for European countries are similar in magnitude. See Sraer and Thesmar (2007) for listed french companies, Bennedsen et al. (2007) for the case of Denmark and Bertrand and Schoar (2006) for South America and Asia.

and Morck et al. (2000)).[3] The underperformance is generally explained by the conjecture that the management ability of the descendent CEOs can be expected to be lower, as the ability of the first choice professional manager.

A complementary explanation for the underperformance of the descendents CEOs is that high ability managers adversely select away from family firms. Strong family affiliations may negatively affect non-family managers in family managed firms. Promotion tournaments and important decisions are likely to be distorted in the favor of the family members. Poza et al. (1997), for example, report, that non-family managers see their career path and chances for senior positions as much worse than family members. Bloom et al. (2011) report that reaching higher management positions in the Indian textile industry is nearly impossible without a family affiliation.[4] Anticipating these distortions, high ability managers may leave family firms or initially select away from them into firms with better career perspectives. Firms with lower average abilities in management or senior management positions are more likely to underperform, compared to non-family firms.

This paper points out adverse selection effects as a possible explanation for firms' underperformance and suggests transparent managerial incentives for limiting the negative effect of strong social ties. To date, these points have been widely neglected in the family firms and favoritism literature.

The experimental design of this investigation builds upon Brandts and Solà (2010), who study the reaction of outsiders on favoritism in organizations. In their design, a manager distributes a pie between a friend and a third unknown player. After the distribution, the recipients give back an arbitrary amount of money. The authors find strong direct favoritism, such that managers favor their friend in the pie sharing game, but do not find evidence for outsiders' "revenge", i.e., disadvantaged players do not return less than non-disadvantaged players. Brandts et al. (2006) use a similar game structure for studying distribution decisions if the distributor is randomly

[3] This result is not totally unchallenged. Anderson and Reeb (2003) as well as Sraer and Thesmar (2007) report positive correlations for some of their performance measures and family CEOs (including descendent CEOs).

[4] For an overview of the relationship between family managers and non-family managers see Chua et al. (2003).

assigned or deliberately chosen by a player.

Bandiera et al. (2009) find that even moderate social ties have an impact on inner firm decisions and production efficiency. In the absence of managerial incentives, fruit pickers with a social connection to the supervising manager have a 5% - 10% higher productivity than unconnected workers. The gap is explained with nepotistic behavior by supervising managers. In the presence of managerial incentives, the productivity gap between connected and unconnected workers vanishes. Managers seem to allocate their supporting effort away from connected workers to high productivity workers. This turns out to increase the total efficiency of production.

This paper mimics the settings in Bandiera et al. (2009) but adds the participation decisions of the unrelated employees. With this experimental design, we study selection effects induced by nepotistic behavior, which has not been analyzed in the literature as of yet.

The next section introduces the experimental design, hypotheses, and experimental procedures. Section 4.3 summarizes the results while section 4.4 eloborates upon the main findings and policy implications.

4.2 Design and Hypotheses

In the experiment, 3-person groups play a 2-stage game. Figure 4.1 shows the game's structure and respective payoffs. In the first stage of the game, Player C decides between entering the second stage with a pie sharing game (choosing G) or taking an outside option (choosing O). The outside option yields 10€ for Player C, 8€ for Player A and 4€ for Player B.[5] If C chooses O, the game ends. If C chooses G, Player A distributes the shares of 3€ and 5€ to Players C and B. The amount sent to Player C is tripled by the experimenter. Choosing action E yields the efficient allocation of 3€ for player B and 15€ for Player C. Choosing I yields the inefficient allocation with 5€ for B and 9€ for C. In the baseline treatment, Player A gets 8€, independent of her distribution decision.

[5]By the time of the experiment, 10€ were equivalent to approximatly US $14.

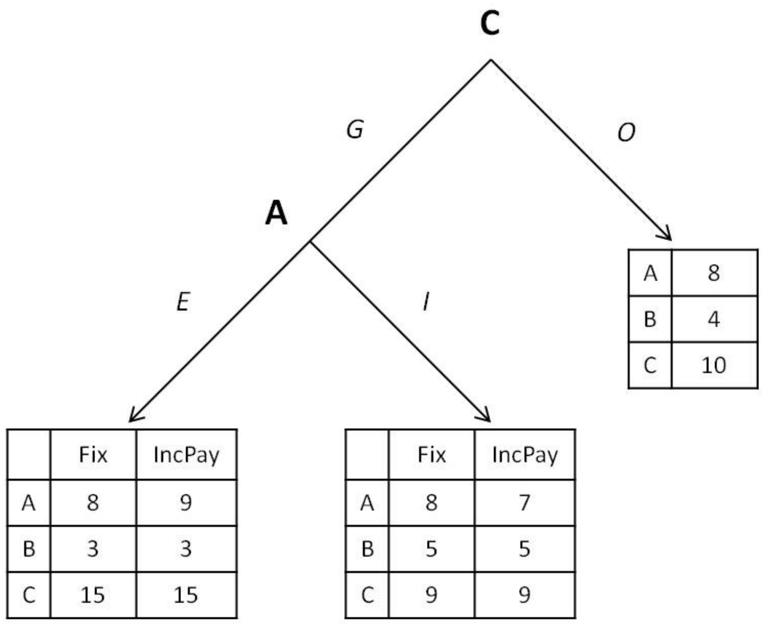

Figure 4.1: A 2-Stage Pie Sharing Game

To study the effect of social ties and incentives, we conduct *Friend* and *Incentive* treatments. In the *Friend* treatments, Players A and B know each other. During the recruitment process, 2-person groups were encouraged to sign up for the experiment. Group members were randomly assigned to a player type (A and B) and matched to an unknown Player C. The recruiting procedure and group constellations were explained to all participants in the respective treatment. In the *Incentive* treatments, Player A does not get a fixed payment for her distribution decision, but receives 50% of the payoffs of Player A and Player B, and thus, profits from choosing allocation *E,* rather then the inefficient allocation *I.* Table 4.1 displays the treatment names of the resulting 2x2 design.

Compared to the outside option, Player C gains 5€ under the efficient allocation and loses 1€ under the inefficient allocation. This insures that Player C has a "natural" inclination to choose *G,* such that social ties in the *Friend* treatments can have a negative effect on C's participation decision.

The experimental design mimics the situation in which a high ability applicant can enter an organization or take an alternative. When entering the firm, she is dependent on the goodwill of her boss (Player A), who has to decide whether to favor employee B or employee C. The multiplicative factor for Player C's payoff can be seen as a higher productivity of Player C. Distributing the larger share to C (B) generates total payments of 26€ (22€) in treatments with no incentive payments and 27€ (21€) in treatments with incentive payments. Thus, favoring Player C always yields the efficient, and hence, desirable allocation from the organization's point of view.

	Compensation Player A	
Social Ties	Fix Pay	Incentive Pay
A and B are friends	*FriendFix*	*FriendInc*
Players are anonymous	*AnonymFix*	*AnonymInc*

Table 4.1: 2x2 Treatment Design

Hypotheses

In the baseline treatment *AnonymFix*, player A neither has a monetary incentive to choose a certain allocation, nor a social inclination for favoring one of the other players. However, in treatment *FriendFix*, Player A is related to Player B, and hence more likely to favor Player B than in the baseline.

> H_1 *(Favoritism): Player A more often chooses action I in FriendFix than in AnonymFix.*

The allocations in the baseline, resulting from actions E and I differ in two respects. First, allocation E is the efficient allocation, since the sum of the players' payments (8€, 3€ and 15€) is 4€ higher than under I. Second, allocation I yields a much more equal distribution, since payments reach from 7€ to 9€. Hence, the decision in the *AnonymFix* treatment is probably driven by equity and/or efficiency concerns. To detect favoritism, not all A Players should choose the inefficient allocation, due to strong equity preferences. If participants would do so, social ties in the *FriendFix* treatment could not increase the frequency of action I, as stated in H_1. The discussion comparing efficiency and inequity motives in allocation decisions suggests that A Player decisions should be sufficiently heterogeneous to make a favoritism effect possible.[6]

In treatment *FriendInc* Player A bears a financial loss when favoring her friend. Choosing E yields a 29% higher income than choosing I.

> *H_2 (Crowding out): Player A more often chooses action E in FriendInc than in FriendFix.*

Hypotheses H_1 and H_2 closely mimic the results in Bandiera et al. (2009), where managers react to a change from a fix to a incentive pay by following economical, rather than social, motives. The predictions regarding Player C's behavior are based on the behavior of Player A in stage 2. Anticipating H_1, Player C is more reluctant to enter the pie sharing game in *FriendFix*.

> *H_3 (Selection effect): Player C more often chooses action G in AnonymFix than in FriendFix.*

Assuming that financial incentives at least partially crowd out A's favoritism, C's participation rate should increase in *FriendInc*.

> *H_4 (Selection crowding out): Player C more often chooses G in FriendInc than in FriendFix.*

Experimental Procedures

We conducted 6 sessions between January and May of 2011 in the Cologne Laboratory for Economic Research in Cologne, Germany. The total number of participants was 117, yielding 39 independent group observations. A session lasted for 50-60 minutes, with the average payoff of 11.91€ being slightly higher than the hourly wage for typical student jobs. The experiment was computer based and designed with zTree (Fischbacher (2007)). Because of a relatively small target sample size, I used the strategy method (Selten

[6]See Engelmann and Strobel (2004), Fehr et al. (2006), Bolton and Ockenfels (2006), and Engelmann and Strobel (2006) for the ambiguous discussion about social and efficiency preferences.

(1967)). To minimize the chances of contact between group participants, the campus was divided into two parts. In one part, we encouraged single individuals, in the other part, 2-person groups to enroll for the experiment. Player roles were assigned randomly in the laboratory. The scenario was explained in neutral terms. You find an English version of the instructions in section 4.5.2.

4.3 Results

Table 4.2 presents Player A's decision, the respective number of group observations and the proportions of efficient allocations for each treatment. The rate of efficient allocation drops from 38% in *AnonymFix* to 0% in *FriendFix.* This difference is significant at the 10% level ($p = 0.07$), and thus, supports the favoritism hypothesis H_1, that in the presence of social ties, Player A has a strong tendency to favor her friend.[7] The *AnonymFix* resembles a treatment in Brandts and Solà (2010), in which Player A has to distribute 4€ and 6€ between her friend and a high ability player who receives 3.25 times of the transferred share. Similar to our results, 92% of all A-players choose the inefficient allocation, and thus, favor their friend.

	AnonymFix	FriendFix	FriendInc	AnonymInc
I	5	10	6	0
E	3	0	7	8
Observations	8	10	13	8
Proportion E	38%	0%	54%	100%

Table 4.2: Decision Player A

In the *FriendInc* treatment, 54% of all A-Players choose the efficient allocation. The difference between *FriendFix* and *FriendInc* is highly significant ($p = 0.007$). This result reveals, that decision-makers react to financial incentives by choosing the payoff maximizing action *E,* as supposed in H_2,

[7]Due to the small sample size, test results are based on the two-tailed Fisher exact test. Applying the Chi^2 test yields always lower p-values. See table 4.4 and table 4.5 for all between-treatments test results.

more often. Interestingly, the results from *AnonymFix* and *FriendInc* do not significantly differ, so adopting financial incentives in a setting with social ties yields a similar decision pattern as fixed payments without social ties. The results from *AnonymInc* fit the pattern. Compared to *FriendInc*, the absence of social ties raises the proportion of efficient allocations from 54% to 100% ($p = 0.026$).

Table 4.3 summarizes Player C's decision. As expected, the participation rate in the baseline is high. Almost 90% of all C-Players enter the game, while only 50% do so in *FriendFix*. The difference between both treatments is not significant ($p = 0.15$), hence, we find no support for H_3. An explanation for the non-finding is the low risk Player C faces when she enters the game. If A chooses the inefficient allocation, Player C only looses 1€, compared to the outside option, but gains 5€ if A chooses *E*. However, A Player decisions yield an (ex post) expected payoff of 9€ for Player C, such that it was optimal not to enter the game, but to choose the outside option. In *FriendInc*, 92% of all C Players choose *G*. The difference between *FriendFix* and *FriendInc* is significant with $p = 0.052$. As stated in H_4, Player C anticipates A's incentive structure, hoping for a payoff maximizing Player A, and consequently, a higher payoff by their own. Player C is right in doing so. The (ex post) expected payoff is 12.24€, and thus, considerably higher than the outside option payoff.

	AnonymFix	*FriendFix*	*FriendInc*	*AnonymInc*
G	7	5	12	6
O	1	5	1	2
Observations	8	10	13	8
Proportion *G*	88%	50%	92%	75%

Table 4.3: Decision Player C

To measure the strength of social relationships between A and B in *Friend* treatments, participants had to tick on a 10 point scale, from "I almost do not know Player A(B)" to "I know Player A(B) very well". Figure 4.2 in the Appendix illustrates the distribution of the answers. 40% of all participants tick 10, 25% ticked 9 and 35% tick 8 or less on the scale. Figure 4.3 shows the distribution of the difference between the answers of two friends. The answers were highly correlated ($\rho = 0.677$ with $p = 0.000$). The data suggest that participants were very familiar with each other. Concerning player A's behavior, one would expect a positive correlation between choosing I and the friendship measure. The data show weak evidence for this conjecture. Correlating A's decision dummy (=1 if A chooses the inefficient allocation and =0 if A chooses the efficient allocation) with the measure of friendship yields a coefficient of correlation of $\rho = 0.334$ with $p = 0.119$, and hence, no significant correlation. For only comparing groups with close and not so close friends, we correlate the upper 40% with the lower 40% of A Player answers. For this, the friendship measure was dichotomized in participants who know Player B very well (ticking 10 on the scale) and who do not know Player B very well (ticking 8 or less on the scale). Correlating the resulting dummy with Player A's decision yields a spearman coefficient of correlation of $\rho = 0.426$ with $p = 0.088$. Hence, A Players who know their friend very well are less likely to choose the efficient allocation, compared to those who do not know their friend very well. Including the answers with 9 on the scale does yield a positive but insignificant correlation. The fact that a correlation coefficient of $\rho = 0.426$ is close to being insignificant illustrates, that significance is hard to obtain with this sample size.

Similar to other experimental settings, where social ties are involved (Abbink et al. (2006) and Brandts and Solà (2010)), the collusion of related players might be an alternative explanation for the results presented above. Since A and B sign up together for the experiment, they could implicitly or explicitly adopt a profit sharing rule such that they equally split the total earnings after the experiment. The sum of the payoffs of Players A and B in the *FriendInc* equals 12€ for both distribution decisions, E and I. Hence, from the monetary point of view and assuming a profit sharing rule, Player A is indifferent between E and I. If Player A has a slight inclination to be nice to Player C and the earnings are split anyway, she will take action E, but not due to their own payoff motives. If Player C adopts this mechanism, she will more frequently participate in the *FriendInc* than in the *AnonymInc* treatment. Following this idea, the presented results were not due to the crowding out of social ties by monetary incentives, but instead were due to the possibility of a joint budget in the *Friend* treatments. Assuming close friends to be more likely to adopt a sharing rule compared to not so close friends, we would expect the following correlation: A Players with close friends should more often choose the efficient allocation, since they are more likely to equally split their total profits than A Players with friends that are not as close. As argued in the former passage, there is some evidence, that A Players who are more closely related to their companion are less likely to choose the efficient allocation. In this light, the collusive behavior of friends seems unlikely.

After the experiment, the participants were asked for basic socioeconomic information and used the Holt-Laury measure of risk for eliciting agents risk preferences (Holt and Laury (2002)). According to the game setting, one would expect more risk averse C Players to be less likely to enter the game with an uncertain payoff. But Players' risk aversions or socioeconomic factors were not found to be correlated with the decision behavior.

4.4 Conclusion

We introduced a 2-stage, 3-person pie sharing game to study the effect of strong social ties on distribution decisions and the participation decisions of unrelated players. C Players had to decide whether to enter a second stage and facing a pie sharing game, or taking an outside option. In the second stage, A Players distributed 3€ and 5€ between C and B Players, where C Players receive the tripled amount sent by A. B Players had no active role. We study two treatment variations. In *Friend* treatments, Player A and B were friends. In *Incentive* treatments, Player A has a monetary incentive for sending 5€ to Player C.

In the *Friend* treatments, decision-makers show favoritism by sending the larger share of the pie to their friends. In the *Incentive* treatments, decision-makers choose more often the efficient distributions. C Players anticipate Player A's decision by proceeding more often to the pie sharing game in the presence of incentives. Thus, in this experimental setting, financial incentives decrease nepotistic behavior and crowd out selection effects.

Applied to real world scenarios, the results discussed above can help to clarify the role of incentives in two ways. First, financial incentives in organizations might reduce nepotistic behavior and lead to more efficient decisions. In real world terms, efficiently distributing a pie could be interpreted as promoting the more able, rather than the more liked, person, as in Berger et al. (2011), or supporting high ability employees, rather than friends, as in (Bandiera et al. (2009)). Second, and most widely neglected so far, proper financial incentives can help to avoid self selection of high ability workers who refrain from entering organizations in which a fair reward to performance is threatened by family ties or strong social ties. If the organization in question manages to design incentive schemes in such a way that outsiders expect a fair reward for their performance, the average quality of hired managers, and hence, the firm performance could rise.

4.5 Appendix to Chapter 4

4.5.1 Tables and Figures

	Decision Player A			
	AnonymFix	FriendFix	FriendInc	AnonymInc
AnonymFix	-	0.069	0.659	0.026
FriendFix	0.069	-	0.007	0.000
FriendInc	0.392	0.007	-	0.046
AnonymInc	0.013	0.000	0.032	-

Table 4.4: Decision Player A: p-values for a two sided (above diagonal) and one sided (below diagonal) Fisher exact Test between Treatments

	Decision Player C			
	AnonymFix	FriendFix	FriendInc	AnonymInc
AnonymFix	-	0.152	1.000	1.000
FriendFix	0.12	-	0.052	0.367
FriendInc	0.629	0.035	-	0.531
AnonymInc	0.5	0.278	0.316	-

Table 4.5: Decision Player C: p-values for a two sided (above diagonal) and one sided (below diagonal) Fisher exact Test between Treatments

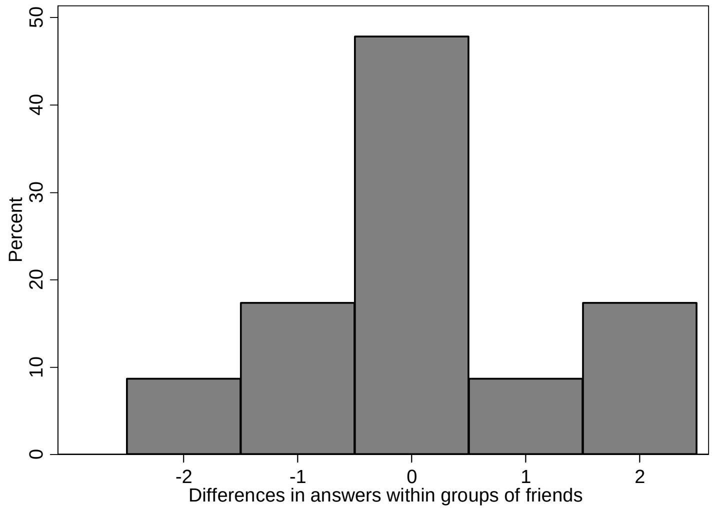

Figure 4.2: The empirical Distribution of the Friendship Measure

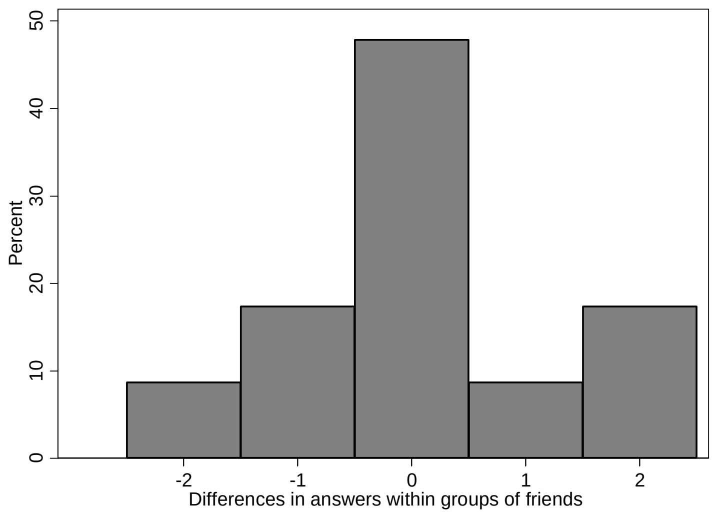

Figure 4.3: The empirical Distribution of Within-Group Differences in Friendship Measures

4.5.2 Instructions

The original instructions are in German. Below you find the English translation. Please contact the author for the original instructions. The following instructions are taken from the *FriendInc* treatment. Deviations from other treatments are commented in italic letters.

Welcome to today's experiment

Please read the following instructions carefully. In the instructions, you will be told everything you need to know for participating in this experiment. If you do not understand something, please raise your hand. Your question will be answered privately at your cabin by one of the experimenter.

In the course of the experiment you can earn money. How much you earn depends on your decisions and the decisions of other participants. At the end of the experiment, the money you earned will be paid out in cash. Please wait in your cabin until the number of your cabin number is called. Please take all the documents you have gotten from us and hand them back to us after the payment. In addition to the payment from the actual experiment, you will get a show-up fee of 2.50€ and 1.50€ for filling out a questionnaire at the end of the experiment.

Please notice that you may not communicate with others during the entire experiment. Additionally, we have to point out that you may only use those functions of your computer which are necessary for the experiment. Communicating or leaving the experiment screen at your computer will lead to your exclusion from the experiment.

If you have any questions, please raise your hand, we will come to your cabin.

Registration of participants

The Registration for this experiment was conducted in two different ways. Some participants were invited to participate alone in an experiment on human behavior. In the other case we invited 2-person groups for participating in an experiment on human behavior. The members of a 2-person group had to register together and therefore know each other. (*The Registration part is missing in the Anonym treatments.*)

The experiment

At the beginning of the experiment, groups are formed which consist of 3 participants. (*In the Anonym treatments we added "randomly"*) In each group there are 2 participants who have registered together: Player A and player B. The third participant, player C, has registered alone for the experiment. (*In the Anonym treatments, the last two sentences are missing. Instead I added: "Every group consists out of a player A, a player B and a player C."*)

Decision player C

Player C makes exactly one decision. He may choose whether he participates in a game with player A and player B or whether he does not participate. If player C decides not to participate, players receive the following payments.

	Payments
Player A	8€
Player B	4€
Player C	10€

The decision of player A does not affect the payment then. If player C decides to participate in the game with player A and Player B, the payment depends on the decision of player A. Player A and player C do not know at the time of their decision how the other player is going to decide. Thus, player A does not know whether Player C will participate in the game. Player C does not know how player A decides.

Decision player A

Player A makes exactly one decision and does not know how player C decides. Player A gets an amount of 8€. Player A now has to decide how he distributes the amount among player B and player C. He has to give 3€ to one player and 5€ to the other player. After player A has distributed the money it is transferred to the other players, and the amount sent to player C is additionally multiplied by 3. Player A's own payment is dependent on the payments of the other players. He gets half the payment of the other two players. (*In the Fix treatments the last sentence is replaced by "Independent of the distribution decision player A receives 8€."*)

The following payments can arise from the decision:

		B receives	A receives	C receives
A sends	3€ to B and 5€ to C	3€	3x5€=15€	0.5x3€+0.5x15€=9€
A sends	5€ to B und 3€ to C	5€	3x3€=9€	0.5x5€+0.5x9€=7€

(*In the Fix treatment the last column shows an 8€ payoff for player A.*)

Whether the decision of player A actually leads to the payment listed in the table depends on whether Player C has decided to participate in the game.

Summary payment

If player C has decided not to participate, each participant gets a payment according to the table on page 2, independent of the decision of player A. If player C has decided to participate in the game, the payments result from player A's decision. (table on this page)

Course of the experiment

Before the experiment, there will be three questions regarding your comprehension. The experiment will only start when each participant has answered the questions correctly. At the beginning of the experiment, it is randomly determined which participant of the registered 2-person groups will be player A and who will be Player B. After the experiment, another instruction will be distributed. Following this, we kindly ask you to fill out a questionnaire.

Please remain seated after you have filled out the questionnaire until we call the number of your cabin. Bring this instruction and the number of your cabin to the front. Only then can the payment of your game results be made.

(*The third sentence of this paragraph is replaced in the Anonym treatments: "At the beginning of the experiment, we randomly form 3-person groups".*)

The decision situation

In this experiment, you have to make 10 decisions. We show you 10 lines on your screen. In each line, you make one decision. Each decision is a choice between an option A and an option B. In both option A and option B you can win a certain amount of money with a probability of x (in %) or another, slightly lower amount with the converse probability 1-x. For example, should you choose line 2 you get 2.00 € with a probability of 20% and 1.60€ with a probability of 80%. Option A and option B only differ in the height of the amount. As soon as all players have made their ten decisions, two lotteries are drawn. The first lottery randomly picks one of the 10 lines. Depending on the Option you chose for this line, either option A or option B gets relevant for your pay. According to the given probabilities, the second lottery decides which amount you actually get, the higher or the lower one. Among all the participants of the experiment we randomly draw 4 persons who get the amount at the end of the as additional pay. If you have questions, raise your hand and we will come to your cabin.

Chapter 5

Result-Based Sampling in Experimental Economics[1]

5.1 Introduction

In experimental economics, a stylized project could run as follows: Researchers develop an experimental design to test an economic hypothesis. They then decide upon the initial sample size, plan and conduct control and treatment group sessions, and examine the hypothesis with a significance test (e.g., t-test, Mann-Whitney U test). The critical level of significance is commonly set at 5% or 10% and hypotheses are only supported if the p-value is below the chosen threshold. Whenever the results are statistically significant or clearly above the significance level, no further data is gathered. However, if the p-value is slightly above the significance level, researchers might decide to increase the initial sample size by conducting additional experimental sessions. The guiding rationale suggests that there might indeed be the hypothesized effect, but a larger sample size is needed to uncover this relationship and obtain statistical significance. That is, the decision whether to increase sample size is contingent on results from the initial sample. This approach, which we label *result-based sampling* (RBS) in this article, is a common practice in experimental economics.

[1] This chapter is based upon Biemann and Herbertz (2011).

However, there is a major problem with RBS that has not been addressed so far in the literature. The significance level determines the probability that researchers find significant results, even if there is no underlying effect. Thus, the higher the significance level, the higher the chance to wrongly find a treatment effect, even when there is no such effect. Findings that wrongly support the existence of an effect are Type I Errors (also known as false positives or alpha errors). To control Type I Error rates, researchers scrupulously validate their empirical results against the common significance levels at 5% or 10%. Whenever p-values are higher than the threshold, results are reported as insignificant to avoid committing Type I Errors. In RBS, researchers first test against the pre-determined significance level. If and only if the result is not significant, additional data is gathered and researchers get a second chance to find significant results, even if there is no true effect. This adds to the initial significance level and the actual significance level will therefore always be higher in RBS than the pre-defined significance level of 5% or 10%. RBS produces an inflation of the Type I Error.

Simmons et al. (2011) show that a step-wise sample size extension increases the Type I Error rate. In addition to Simmons et al. (2011), we explore how, and to which extent RBS causes Type I Error inflation in experimental economics.[2] We examine whether this problem of RBS is practically relevant to researchers, as opposed to being merely an exercise in statistical pedantry. By simulating settings commonly encountered in experimental economics, we provide evidence of a substantial Type I Error inflation. Under reasonable assumptions, RBS results in 7.7% wrong positive results if the chosen level of significance is 5%.[3] Consequently, we provide suggestions for how to overcome this inflation.

The remainder of this article is organized as follows. In section 5.2, we briefly discuss Type I Error inflation in statistical hypothesis testing by pointing out similarities and differences between RBS and multiple testing in general. In section 5.3, we introduce our simulation approach and analyze the

[2] Note, that by the time of our simulation (May to September 2011) Simmons et al. (2011) was not published. A working paper was not available.

[3] As will be explained in section 5.4, these numbers stem from scenarios with a maximum number of 2 RBSs and a RBS threshold of 20%.

impact of RBS on Type I Error. Section 5.4 examines important determinants of Type I Error inflation in RBS. Lastly, in section 5.5 and section 5.6, we discuss our results and make suggestions for how to overcome or attenuate Type I Error inflation in RBS. Section 5.7 concludes.

5.2 Type I Error Inflation in Statistical Hypothesis Testing

RBS involves the conditional application of more than one statistical test on related datasets. If the first test fails, the dataset is extended and a second test is performed. For a single statistical test, the probability of making a Type I Error is correctly indicated by the significance level. Whenever a multitude of tests are applied to a dataset, the probability of making a Type I Error for this family of tests is higher than for the single test. We are not aware of any literature that discusses this problem for RBS. We will therefore ground our work on a discussion for group comparisons that involve more than two groups (e.g., ANOVA). Methodological considerations from this related area can help to gain a better understanding of the problem in RBS. Similarities and differences to RBS will later be used to develop simulation settings and discuss statistical remedies for RBS.

Multiple group comparisons are frequently encountered in experimental economics whenever several experimental conditions are assessed. Let us assume that mean differences between a control group and five independent treatment groups are compared. The Type I Error rate is set to 5% for each test, but the probability that one or more of these tests yields a statistically significant result is $1 - 0.95^5 = 0.226$. The importance of looking at this much higher familywise error rate is well-known for multiple comparisons. Several procedures have been suggested in the literature that correct for the increased Type I Error. For example, when a Bonferroni correction in ANOVA is applied, a result from a pair-wise comparison is considered statistically significant if the p-value is smaller than $alpha/n$, where n is the number of independent tests performed on the data and $alpha$ denotes the chosen significance level. When a control group is compared to five treatment groups and alpha = 5%, the adjusted significance level is 0.05/5 = 0.01. Thus, the p-value of any comparison must be smaller than 1% to be interpreted as a statistically significant result at the 5% level. Other procedures are the LSD test, Scheffé method, Tukey's HSD, and the Šidák correction (Savin (1980)).

These multiple comparison procedures were designed for tests that are conducted simultaneously and independently. In RBS, a second test is only performed if the previous result was non-significant. Furthermore, results from additional tests in RBS are not independent from the previous test, as the initial sample is extended, but not replaced. That is, all observations from the initial sample are part of the extended sample. This follows that a familywise error rate cannot be computed, as in the case of n independent tests, which, in turn, makes it impossible to transfer existing multiple comparison procedures to RBS. The most important difference between RBS and multiple testing is the sequential approach in RBS, because the application of further statistical tests is contingent on the results of the previous test. When researchers use an uncorrected significance level in the initial sample (i.e., 5% or 10%), any finding that turns a non-significant finding from the first trial into a significant finding in extended samples would add to the initial error rate. It follows that Type I Error inflation can only be avoided if *all* researchers test their *initial* sample against a significance level below the common significance level of a single test, keeping in mind that they would increase the sample size in the case of the first test being non-significant. This procedure is impossible from the practical point of view. However, the extent to which Type I Errors are inflated has yet to be demonstrated in the literature. With this goal in mind, we report results from Monte Carlo simulations in the next section. Subsequently, we will use simulation results to discuss means to avoid or alleviate Type I Error inflation in RBS.

5.3 Simulation and Results

We conceptualized RBS as a stepwise approach that starts with an initial sample that is extended, if the first results are non-significant. This sequence of conditional events is reflected in the simulation settings. To illustrate our simulation approach, we have chosen a simple and coherent example, later generalized to other settings. Let us assume that researchers state an economic hypothesis in that the mean value in a treatment group A is not equal to the mean value in the control group B, hence $\mu_A \neq \mu_B$. They conduct an experiment in period 1 and gain a total of 40 observations that are equally split over treatment and control group ($n_{experimental} = n_{control} = 20$). They then test the null hypothesis $H_0 : \mu_A = \mu_B$ with a t-test and a significance level of 5% (two-tailed). We simulated this experiment by drawing 20 observations for group A and 20 observations for group B from a normally distributed variable (M = 0, s.d. = 1). Our goal was an estimation of Type I Error inflation. This error only occurs if there is no true effect, but significant results wrongly indicate the existence of an effect. The absence of a true effect is simulated via identical mean values for experimental and control group ($M_{experimental} = M_{control} = 0$). The comparison of group means with a statistical test can fall into one of three categories (see figure 5.1):

$p < 5\%$. The researchers find support for the alternative hypothesis $\mu_A \neq \mu_B$. They conclude that no further sessions are needed and aim to publish the results.

$5\% < p < p_{RBS}$. Results are not significant, but the p-value is so low that it seems appropriate to increase the sample size, i.e., apply RBS. This upper limit of the p-value for RBS is denoted by p_{RBS}. The underlying assumption is that researchers will not necessarily increase the sample size. Instead, they will evaluate costs and potential benefits of an extended sample. We therefore defined the critical p-value p_{RBS}, which reflects these considerations. If the p-value from the first statistical test is within this range, additional data are gathered to increase the sample size. We implemented this step in the simulation by adding 20 observations to the initial sample size in an additional period, again being equally split over treatment and control group. A

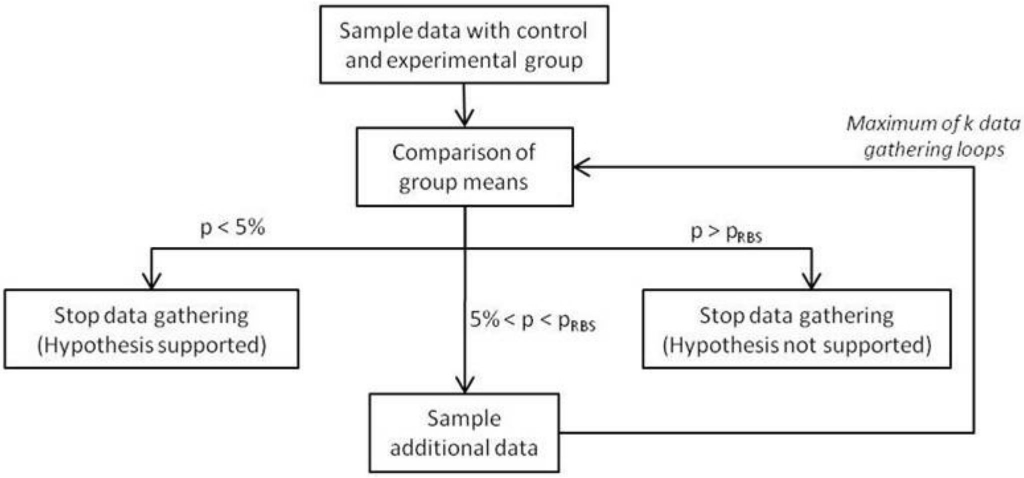

Figure 5.1: The Simulation Process

t-test was then calculated that used the total sample.

$p > p_{RBS}$. The p-values are above the threshold p_{RBS} that would justify additional data gathering. Similar to the first case with a significant result, researchers stop the data gathering process at this point, but will most likely not try to publish the result, as they are aware of the connection between significant results and publication probability (Sterling (1959) and Easterbrook et al. (1991)).

Overall, the simulation procedure, as depicted in figure 5.1, contains a conditional loop executed whenever the p-value of the sample is non-significant, but is below p_{RBS}. As a further limitation, we restricted the maximum number of additional samples to k, because it is not plausible to assume that researchers extend the sample an unlimited number of times. In the most simple case, k is set to one, which means that researchers might increase the initial sample a single time. Results from this setting are illustrated in figure 5.2.[4] Figure 5.2 shows a histogram of p-values for the first and second period (i.e., k = 1) with $p_{RBS} < 20\%$ and $M_{experimental} = M_{control}$. The upper part of the figure illustrates the simulated p-values from the first

[4] All simulations were coded in the R language of statistical computing (R-Development-Team (2008)) based on 1,000,000 drawings for each setting.

period. Because there is no true underlying effect, p-values follow a uniform distribution and, for example, 4.9% of all simulated p-values are found in the interval from 0% to 5%. Thus, in 4.9% of all simulated experiments, the result is statistically significant, although there is no true effect. This correctly represents the Type I Error rate of 5% with a small deviation due to sampling error. In period 2, we simulated additional data whenever the results in period 1 were non-significant, but below the threshold of 20% ($p_{RBS} < 20\%$). This was the case in about 15% of all simulated experiments. The lower part of figure 5.2 shows results after period 2. Most importantly, the percentage of findings with a p-value below 5% increased from 4.9% in period 1 to 6.8% in period 2. Thus, from the 15% of samples that were extended in period 2, about 1.9% delivered statistically significant result. Simulation parameters were chosen to show no true effect, and hence, all significant results were Type I Errors. The probability of committing a Type I Error was therefore not 5% as indicated, but about 6.9% instead; a relative increase of about 39% percent. In other words, researchers that apply RBS have a 39% higher chance of wrongly finding statistical differences in settings where no true effect is present. In the following, we will refer to this actual fraction of wrong positive results as α_{RBS}. The difference between the "true" significance level of 5% or 10% and α_{RBS} defines the inflation of the Type I Error due to RBS.

5.4 Determinants of Type I Error Inflation

The simulation in the previous section was restricted to a specific set of parameters. We now present results from simulations in which we systematically vary the most important parameters. More specifically, we analyze the Type I Error inflation when the upper limit of p-values (p_{RBS}) for additional data changes, the maximum number of additional data gathering periods (k) is altered, and other statistical tests are applied.

Figure 5.3 graphically depicts simulation results when k and p_{RBS} are varied. Table 5.1 in the Appendix illustrates the respective numerical simulation results. The horizontal rhs axis in figure 5.3 plots k, the horizontal lhs axis plots the respective threshold p_{RBS} in 2.5% steps, and the vertical axis

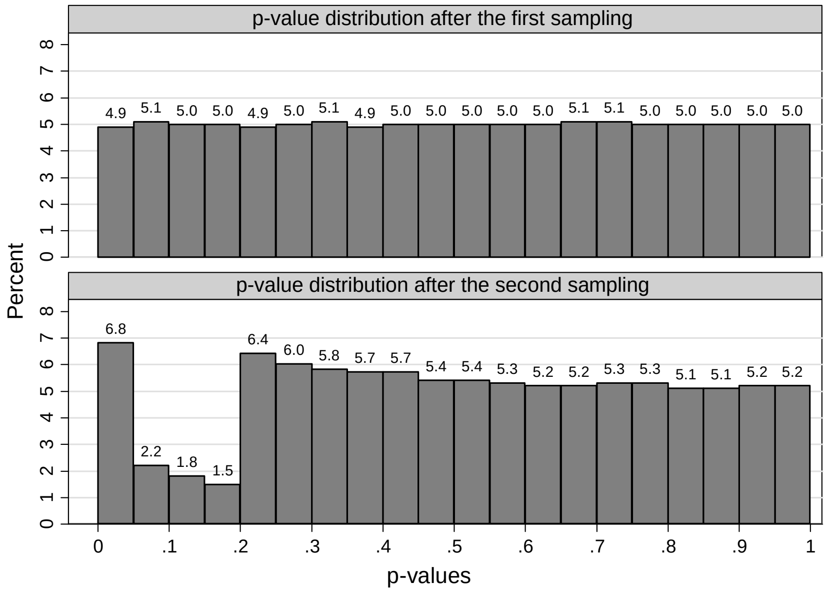

Figure 5.2: Histogram of p-values after first and second Sampling

shows the actual alpha error α_{RBS}, which is estimated from the simulations. As can be seen in the graph, the actual Type I Error increases with both k and p_{RBS}. When k increases, there were more periods in which samples in the critical area of p-values ($5\% < p < p_{RBS}$) were extended and got another chance to reach statistical significance. Furthermore, figure 5.3 indicates that Type I Error inflation decreases with k, because in each period, fewer samples were left in the critical area of p-values, and hence, less samples were supplemented with additional data. As can be seen from the upper panel in figure 5.2, about 15% of all samples had p-values in the $5\% < p < 20\%$ interval after period 1 and only 5.5% after period 2. Thus, only 5.5% are found in the critical range in which hypothetical researchers would add additional data in a further period . Hence, the Type I Error inflation effect is decreasing with k.

With an increasing p_{RBS}, the number of samples that were extended in each period was higher and, hence, there were more samples that became significant after adding data. This is shown by the positive trend of α_{RBS}

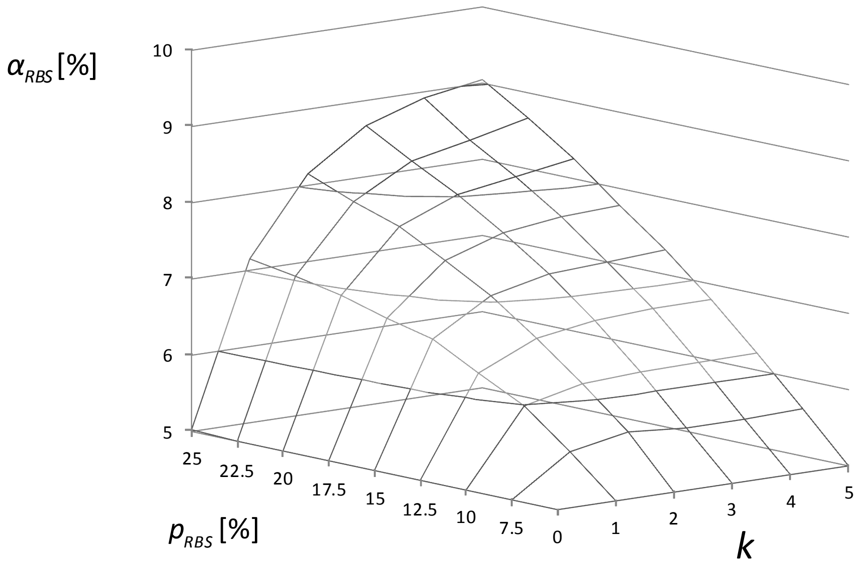

Figure 5.3: α_{RBS} in Dependence of the Sampling Threshold p_{RBS} and the Number of Samplings k

with increasing p_{RBS}. Again, the additional effect on α_{RBS} is decreasing with an increase of p_{RBS}. The reason is that researchers with insignificant results but relatively low p-values are more likely to get significant results after adding data in the following period than researchers with higher p-values. For example, the Type I Error inflation from raising the threshold p_{RBS} from 10% to 15% is higher than from raising the threshold from 20% to 25%.

In the previous simulations, we drew observations from a normal distribution and conducted t-tests. This constellation is appropriate in the sense that the applied test fits perfectly with the true distribution of the underlying data. Of course, this setting is rare in experimental economics, since non-parametric statistical tests are mostly applied. We conducted a series of additional simulations to test the robustness of our results (details are available from the authors upon request). Overall, we found that our results were robust to variations in the underlying distributions and tests. For example, figure 5.4 and table 5.2 in the Appendix show results for drawing from a normal distribution, but testing with a Mann-Whitney U test. Changes in α_{RBS} are very similar to what we reported in table 5.1 for t-tests. That is not very surprising, as the mechanism that results in an inflation of Type I Errors is a property of RBS itself and not limited to a specific statistical test or distribution.

5.5 Discussion

The goal of this study was to assess Type I Error inflation caused by researchers' strategy to increase an initial sample in an experiment if the results from statistical tests are somewhat above the threshold for statistical significance. By means of simulation, we illustrate that this approach, labeled RBS, yields a substantial underestimation of the true significance level. For example, if the significance level was set to 5% and we further assume that researchers increase the initial data up to two more times ($k = 2$) whenever p-values are between 5% and 20%, the true alpha level is 7.7%. That is, there is a probability of 7.7% (not 5%) that a result is significant, even if there is no true underlying effect. Given the strictness that researchers exhibit when evaluating results of statistical tests, this is an alarming inflation of the Type I Error. Therefore, our conclusion is that experimenters should not apply RBS. However, it must be noted that our recommendation only affects RBS in statistical hypothesis testing. We do not intend to condemn all forms of pre-testing and stepwise sampling. Instead, we would like to offer the following set of guidelines that experimental researchers can use whenever they seek to gather data in a stepwise fashion.

1. Pre-determine sample sizes.

RBS can be avoided if researchers define the target sample size a priori. Often researchers apply RBS because they are unsure whether the initial sample will be of a sufficient size. Power analyses have been developed to avoid these problems. Whenever information is available on expected effect sizes, they have been proven to be a useful tool for a priori sample size calculations. We will discuss this tool in section 5.6.

2. Stay the course in the case of significant pre-results.

Researchers can examine first results while the experiment is running. But results from this study show that researchers should not stop data gathering because significant results were found before the final sample size was reached. It is less critical if researchers stop the data gathering process when results are non-significant, because then Type I Errors cannot be committed. Therefore, we suggest that researchers should only intend to publish research findings when they are based on the full pre-defined sample.

3. Do not extend the pre-defined sample size.

If the pre-defined sample size is reached and results are non-significant, researchers should not continue to add data to the existing dataset. Otherwise, they create the same Type I Error inflation described in this study.

4. Report the sampling approach.

Researchers should report the pre-defined sample size and explain deviations from the actual sample size. There might be situations that justify changes in the sampling procedure. However, whenever adjustments are based on RBS, this should be noted and explained.

5.6 Power Analysis

Problems with RBS let us conclude that researchers must define their sample size before data gathering, instead of extending the sample post-hoc. However, researchers might be unsure whether a pre-defined sample size will be sufficient to test their hypotheses. For researchers, it is important to minimize the probability that a true effect remains undetected in an experiment. This beta error (β) occurs whenever there is a true underlying effect, but the result of a statistical test is not significant and, thus, a correct hypothesis does not find support in the data (Cohen (1988); Murphy and Myors (2004)). The reverse probability (i.e., $1 - \beta$) is the statistical power of a test. It is defined as the probability to reject a null hypothesis when the null hypothesis is indeed false. Although there is vast amount of literature on statistical power and power analyses in the psychological literature and some related fields, little is known about this concept in experimental economics. We will therefore provide a brief overview on power analyses to arm experimental economists with the relevant knowledge to perform ex ante calculations of sufficient sample size.[5]

In statistical inference testing, significance level, population effect size, sample size, and statistical power form a closed system. Any parameter can be computed if values for the other three are set. Thus, the necessary sample size for an experiment can be derived from desired values for significance level and statistical power, and the hypothesized effect size. The significance level is mostly set to 5%. Although there is no convention, a minimum statistical power of 0.80 is considered adequate (Cohen (1988)). That is, the probability that a true effect is not significant in a statistical test is less than 20 percent. Lastly, the effect size must be estimated, because a smaller sample is necessary if the underlying effect is large. Small effects, on the contrary, only have a high chance of being statistically significant in very large samples. For comparisons of mean values, the effect size is expressed in d, defined as the difference of the group means divided by the standard deviation. As a rule of thumb, one can think of large effects whenever d $\geq$ 0.80, medium effects for d around 0.50, and small effects for d around 0.20 (Cohen (1988)). Effect size measures can be found for other statistical tests as well, for example, correlation coefficients and ANOVAs. There are several tools available that perform power calculations, most notably the freeware G*Power (Faul et al. (2007)), which computes the value of the fourth variable, if the other three variables are defined. For example, we are interested in an economic effect with a large effect size (d = 0.80), with a significance level of 5 percent (two-tailed) and a power of 0.80. Then, results from power analyses show that we would need 26 individuals in each of the two groups to have an 80 percent probability of detecting a true effect. For a small effect (d

[5] An extensive discussion of statistical power goes well beyond the scope of this paper. We refer the interested reader to Cohen (1988) and Murphy and Myors (2004).

= 0.20) and similar settings, we would need 394 individuals in each group. Experimental economists would be well advised if power analyses are part of research planning to optimize resource allocation, because large samples may not be necessary to detect large effects and small samples suffer from low statistical power when effect sizes are small. In the latter case, there is a high probability that true small effects are undetected, because they are not likely to be statistically significant in hypothesis testing.

5.7 Conclusion

We show, by means of Monte Carlo simulations, that the practice of RBS in experimental economics leads to substantial Type I Error Inflation. Furthermore, RBS cannot be corrected ex post, because if the hypothesis test with the initial sample is tested against the common thresholds for statistical significance (5% or 10%), any significant finding in extended samples raises the Type I Error above the pre-defined level. Although RBS was shown to bias results from statistical hypothesis tests, for researchers, there is no incentive to avoid RBS. Therefore, we hope that our findings help to create an awareness of the problems associated with RBS, which might then improve research and policy implications in experimental economics.

5.8 Appendix to Chapter 5

	0	1	2	3	4	5
5.0%	0.050	0.050	0.050	0.050	0.050	0.050
7.5%	0.050	0.055	0.057	0.056	0.056	0.056
10.0%	0.050	0.060	0.0612	0.062	0.062	0.062
12.5%	0.050	0.063	0.066	0.068	0.0680	0.0680
15.0%	0.050	0.066	0.071	0.072	0.073	0.073
17.5%	0.050	0.068	0.074	0.076	0.077	0.078
20.0%	0.049	0.069	0.077	0.080	0.081	0.083
22.5%	0.049	0.071	0.079	0.083	0.085	0.087
25.0%	0.050	0.072	0.082	0.087	0.089	0.090

Table 5.1: Simulated Type I Errors from Figure 5.3

	0	1	2	3	4	5
5.0%	0.049	0.049	0.049	0.049	0.049	0.049
7.5%	0.049	0.054	0.055	0.055	0.055	0.055
10.0%	0.048	0.058	0.060	0.061	0.061	0.062
12.5%	0.049	0.061	0.065	0.067	0.066	0.066
15.0%	0.049	0.065	0.070	0.071	0.071	0.072
17.5%	0.049	0.067	0.073	0.075	0.076	0.077
20.0%	0.049	0.069	0.075	0.078	0.080	0.080
22.5%	0.049	0.070	0.078	0.082	0.085	0.085
25.0%	0.049	0.071	0.080	0.084	0.0870	0.088

Table 5.2: Simulated Type I Errors from Figure 5.4

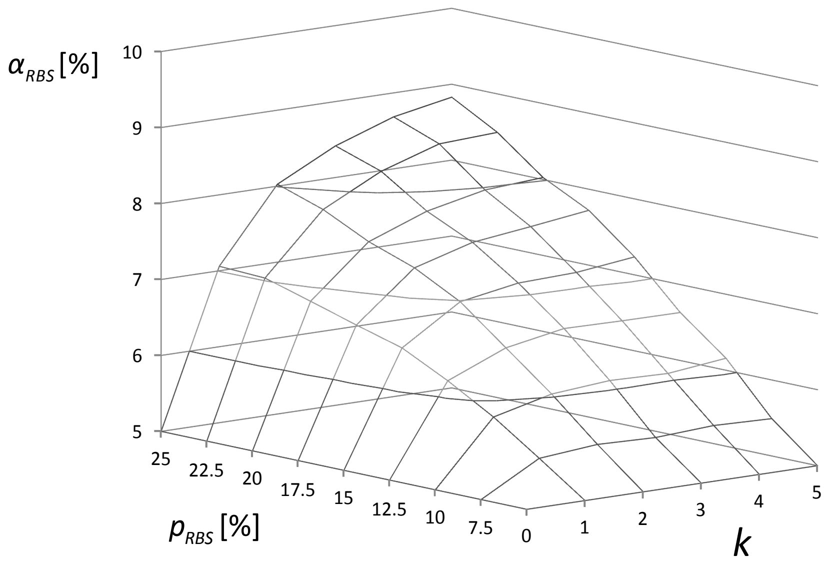

Figure 5.4: α_{RBS} in Dependence of the Sampling Threshold p_{RBS} and the Number of Samplings k for a Mann-Whitney U Test

Bibliography

Abbink, K., B. Irlenbusch, and E. Renner (2006). Group size and social ties in microfinance institutions. *Economic Inquiry 44*, 614–628.

Anderson, R. C. and D. M. Reeb (2003). Founding-family ownership and firm performance: Evidence from the s&p 500. *Journal of Finance 58*, 1301–1327.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics.* Princeton University Press.

Arya, A. and B. Mittendorf (2004). Using job rotation to extract employee information. *Journal of Law, Economics, & Organization 20*, 400–414.

Baiman, S. and M. V. Rajan (1995). The informational advantages of discretionary bonus schemes. *Accounting Review 70*, 557–579.

Baker, G., R. Gibbons, and K. J. Murphy (1994). Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics 109*, 1125–56.

Bandiera, O., I. Barankay, and I. Rasul (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica 77*, 1047–1094.

Bennedsen, M., K. M. Nielsen, F. Perez-Gonzalez, and D. Wolfenzon (2007). Inside the familiy firm: The role of families in succession decisions and performance. *Quarterly Journal of Economics 122*, 647–690.

Berger, J., C. Herbertz, and D. Sliwka (2011). Managerial incentives and favoritism in promotion decisions: Theory and field evidence. *IZA Discussion Paper No. 5543*.

Bertrand, M. and A. Schoar (2006). The role of family in family firms. *Journal of Economic Perspectives 20*, 73–96.

Bhattacharya, S. and J. L. Guasch (1988). Heterogeneity, tournaments, and hierarchies. *Journal of Political Economy 96*, 867–881.

Biemann, T. and C. Herbertz (2011). Result-based sampling in experimental economics. *Mimeo*.

Bjerke, R., J. Cleveland, R. Morrison, and W. Wilson (1987). Officer fitness report evaluation study. *Report TR-88-4 (Navy Personnel Research and Development Center. Washingtion DC)*.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2011). Does management matter? evidence from india. *NBER* (16658).

Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics 122*, 1351–1408.

Bloom, N. and J. Van Reenen (2010). New approaches to surveying organizations. *American Economic Review 100*, 105–109.

Bolton, G. E. and A. Ockenfels (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *American Economic Review 96*, 1906–1911.

Bolton, P. and M. Dewatripont (2005). *Contract Theory.* The MIT Press.

Brandts, J., W. Güth, and A. Stiehler (2006). I want you! an experiment studying motivational effects when assigning distributive power. *Labour Economics 13*, 1 – 17.

Brandts, J. and C. Solà (2010). Personal relations and their effect on behavior in an organizational setting: An experimental study. *Journal of Economic Behavior & Organization 73*, 246–253.

Breuer, K., P. Nieken, and D. Sliwka (2010). Social ties and subjective performance evaluations: An empirical investigation. *IZA Discussion Paper No. 4913*.

Bull, C., A. Schotter, and K. Weigelt (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy 95*, 1–33.

Chua, J. H., J. J. Chrisman, and P. Sharma (2003). Succession and nonsuccession concerns of family firms and agency relationship with nonfamily managers. *Family Business Review 16*, 89–107.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Lawrence Erlbaum Associates.

DeGroot, M. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

Easterbrook, P., R. Gopalan, J. Berlin, and D. Matthews (1991). Publication bias in clinical research. *The Lancet 337*, 867–872.

Engelmann, D. and M. Strobel (2004). Inequality aversion, efficiency and maximin preferences in simple distribution experiments. *American Economic Review 94*, 857–869.

Engelmann, D. and M. Strobel (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Reply. *American Economic Review 96*, 1918–1923.

Eriksson, T. and J. Ortega (2006). The adoption of job rotation: Testing the theories. *Industrial and Labor Relations Review 59*, 653–666.

Fairburn, J. and J. Malcomson (2001). Performance, promotion, and the peter principle. *Review of Economic Studies 68*, 45–66.

Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods 39*, 175–191.

Fehr, E., M. Naef, and K. M. Schmidt (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *American Economic Review 96*, 1912–1917.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*, 171–178.

Granovetter, M. (2005). Impact of social structure on economic outcomes. *The Journal of Economic Perspectives 19*, 33–50.

Grossman, S. and O. Hart (1983). An analysis of the principal-agent problem. *Econometrica 51*, 7–45.

Hauser, F., A. Schubert, and M. Aicher (2008). Unternehmenskultur, arbeitsqualität und mitarbeiterengagement in den unternehmen in deutschland. *Abschlussbericht Forschungsprojekt Nr. 18/05*.

Herbertz, C. (2011). Social ties, incentives, and adverse selection. *Mimeo*.

Herbertz, C. and D. Sliwka (2011). When higher prizes lead to lower efforts - the impact of favoritism in tournament. *Mimeo*.

Holmström, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica 55*, 303–28.

Holmström, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership and job design. *Journal of Law, Economics, & Organization 7*, 24–52.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*, 1644–1655.

Ittner, C. D., D. F. Larcker, and M. W. Meyer (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *Accounting Review 78*, 725–758.

Jensen, Michael C. & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics 3*, 305–360.

Kramarz, F. and O. N. Skans (2007). With a little help from my parents: Family networks and youth labor market entry. *Crest Working Paper*.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review 90*, 1346–62.

Lazear, E. P. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy 89*, 841–864.

Longenecker, C. O., H. P. Sims, and D. A. Gioia (1987). Behind the mask: The politics of employee appraisal. *Academy of Management Executive 1*, 183–193.

Marsh, R. M. (1960). Bureaucratic constraints on nepotism in the ch'ing period. *Journal of Asian Studies 19*, 117–133.

Morck, R. K., D. A. Stangeland, and B. Yeung (2000). Inherited wealth corporate control, and economic growth: The canadian disease? In R. K. Morck (Ed.), *Concentrated Corporate Ownership*. University of Chicaco Press.

Murphy, K. R. and J. N. Cleveland (1995). *Understanding Performance Appraisal*. Thousand Oaks: Sage.

Murphy, K. R. and B. Myors (2004). *Statistical Power Analysis: A Simple General Model for Traditional and Modern Hypothesis Tests.* Lawrence Erlbaum Associates.

Ortega, J. (2001). Job rotation as a learning mechanism. *Management Science 47*, 1361–1370.

Poza, E. P., T. Alfred, and A. Maheshwari (1997). Stakeholder perceptions of culture and management practices in family and family firms - a preliminary report. *Family Business Review 10*, 135–155.

Prendergast, C. and R. Topel (1996). Favoritism in organizations. *Journal of Political Economy 104*, 958–978.

Prendergast, C. J. (1999). The provision of incentives in firms. *Journal of Economic Literature 37*, 7–63.

Prendergast, C. J. (2002). Uncertainty and incentives. *Journal of Labor Economics 20*, 115–37.

Prendergast, C. J. and R. H. Topel (1993). Discretion and bias in performance evaluation. *European Economic Review 37*, 355–65.

Pérez-González, F. (2006). Inherited control and firm performance. *American Economic Review 96*, 1559–1588.

R-Development-Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna Austria ISBN 3.*

Rauch, J. E. and A. Casella (2001). *Networks and Markets.* New York: Russel Sage.

Savin, N. E. (1980). The bonferroni and the scheffé multiple comparison procedures. *Review of Economic Studies 47*, 255–273.

Schotter, A. and K. Weigelt (1992). Asymmetric tournaments, equal opportunity laws, and affirmative action: some experimental results. *Quarterly Journal of Economics 107*, 511–539.

Selten, R. (1967). *Die Strategiemethode zur Erforschung des eingeschränkten rationalen Verhaltens im Rahmen eines Oligopolexperiments*, pp. 136–168. J.C.B. Mohr.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. *Review of Economic Studies 71*, 513–534.

Simmons, J. S., L. D. Nelson, and U. Simonsohn (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science 22*, 1359–1366.

Sraer, D. and D. Thesmar (2007). Performance and behavior of family firms: Evidence from the french stock market. *Journal of the European Economic Association 5*, 709–751.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association 54*, 30–34.

van Dijk, F., J. Sonnemanns, and F. van Winden (2001). Incentive systems in a real effort experiment. *European Economic Review 45*, 187–214.

Vilallonga, B. and R. Amit (2006). How do family ownership, control and management affect firm value? *Journal of Financial Economics 80*, 385–417.