Evaluation Frequency in Employment Relations

Inauguraldissertation
zur Erlangung des akademischen Grades
Doktor rerum politicarum
der Universität Dortmund

Julia Angerhausen

September 2007

Contents

List of Figures

List of Tables

Preface

This dissertation draws on research I undertook during the three years I held a scholarship at the Graduiertenkolleg "Allokationstheorie, Wirtschaftspolitik und kollektive Entscheidungen" and later while I was a teaching and research assistant at the Chair of Microeconomics, both at the University of Dortmund. The Graduiertenkolleg has been financed by the Deutsche Forschungsgemeinschaft (DFG). The present thesis has strongly been influenced by and profited from discussions with professors and fellow students of the Graduiertenkolleg and from presentations given at the Graduiertenkolleg's workshop and at various conferences. I am very grateful to all who supported my work in that way. In particular, I would like to thank Wolfgang Leininger who supervised my dissertation. Moreover, I would like to thank Peter Witt, Burkhard Hehenkamp, Kornelius Kraft, Christian Bayer, Christiane Schuppert, and Frauke Eckermann who all helped me to improve this thesis at the various stages of its development. I would also like to thank the Deutsche Forschungsgemeinschaft for financial support.

Julia Angerhausen

Chapter 1

Introduction

1.1 Incentives, Information, and Bias

Incentives matter! This is a fundamental belief of economists. The aim of this work is to study the design of incentives with respect to timing decisions. More precisely, we will discuss how the frequency with which performance information is retained—i.e. the frequency of performance evaluations—impacts incentives and efficiency within an organization.

Understanding incentives is a major focus of economic theory. In particular, economists are interested in the relation of incentives and efficiency, the question whether incentives can be designed so that the outcome of economic behavior is efficient. An important subject for such kind of investigation are incentives provided to employees via their labor contract and these contracts' properties concerning efficiency. In particular, a large literature deals with contracts comprehending a variable compensation component which is supposed to provide incentives by a linkage of an employee's merit to his pay. This research is highly relevant, as we have observed a continuous rise in the use of variable compensation in the past decades.[1] Hence, understanding how this development impacts the provision of incentives in firms is an important field of economic research.

But the link between pay and performance is less obvious, one could also say less direct, than we might expect initially. What is performance? Can it be measured at all, and if yes, *how* can it be measured? The link between an employee's job performance and his payroll consists in information. The information used to determine to what extent an individual's compensation depends on his performance is pivotal, certainly for the incentives provided by variable pay!

The availability of appropriate information, i.e. an adequate measure, which is able to capture the true value of the work accomplished by an employee differs according to professions. For some jobs, a strong link between performance and some quantitative measure seems natural. This is why we observe, for example, the compensation of salesmen or executives varying largely with the amount of products sold and the performance of the firm respectively. Both of these measures can be easily scaled and quantified by the observation of sales, and a firm's profit or its share price. Additionally, they seem to be highly correlated with performance and, hence, linking compensation to these *objective* measures provides strong incentives to employees in these particular professions.

However, the relation between performance itself and performance *measures* is not always so close. Taking the production process as an example, the number of items produced could be taken as a measure for a worker's performance. Still, performance is not only reflected by the number of items produced but as well by the quality of these products. Furthermore, one would expect that it plays a role for performance whether a worker correctly maintains his workplace, whether he helps to enhance the production process, and so forth. If now, the employer

[1]For example, Parviainen (2003) reports for Finnish industrial firms that the share of blue-collar workers being paid for performance rose from 12.4 to 26.3 percent between 1996 and 2000. The respective share of white-collar workers increased from 28.7 to 37.2 percent in this period. Also, Kurdelbusch (2002) states a rise of variable pay in large German companies since the end of the nineties. This can be seen as the arrival of a trend from Anglo-Saxon countries such as it has been identified in Lebow et al. (1999).

decides to pay this worker only on the basis of the number of produced items, this can lead to a severe distortion of incentives. Performance pay might significantly raise output, but it is quite likely that the worker will at the same time spend less time with quality assurance, maintenance, and process optimization. Activities that he is not rewarded for might be cut back to a detrimental extent, potentially lowering the worker's overall performance instead of boosting it by the recently set incentive pay.

One way to avoid negative effects of distorted incentives based on objective but insufficient measures is to include a subjective component when it comes to the determination of variable pay, e.g. of the amount of a bonus. A lead worker can be additionally assigned the role of supervising the co-workers at his assembly line. When the decision on the amount of the workers' bonus is taken, it can be discussed with this supervisor whether additional "soft" criteria besides increasing production volume have been responsibly respected by the workers. Only when this is the case, the full bonus will be paid.

The more complex the task of an employee is, the more difficult it becomes to find appropriate objective measures for performance. One response to that can be to pay a straight salary instead. Another one is to organize a well-managed evaluation process in which employee performance will be approximated by gathering objective measures as well as *subjective* measures provided by a supervisor or superior who has good insight into the particular employee's work. However, an inherent problem of the use of subjective measures in an evaluation procedure is that they are not contractible and, therefore, the employment relationship must be subject to a certain amount of trust.

So, when an employee's compensation is made contingent on measures that cannot be included in a binding contract, the incentive effect of such a compensation formula crucially depends on the trust between employer and employee, in particular on the employer's reputation concerning the payment of promised bonuses. Another factor that will play a role for incentive provision is the correctness of the information gathered by the means of a subjective evaluation process. Beyond an intentional falsification of the subjective evaluation, a superior might simply make mistakes in his judgments of an employee and, hence, wrongfully lower the agreed bonus. If such behavior is anticipated by an employee, his motivation through variable compensation can be dramatically weakened.

One can think of numerous sources of subjectivity bias in the judgment of a supervisor when this judgment is not based on openly observable and verifiable measures. The bias could be caused by a framing effect, i.e. the judgment will be influenced by former judgments that have been made in prior evaluations. Emotions play an important role in judgments, even when intentional falsifications are omitted. This work will refer to a quite omnipresent human imperfection as a source of supervisor bias, namely the bounded capacity of human memory. As memory decreases over time, the frequency with which employee evaluations are conducted has an impact on the quality of the underlying information.

Employee evaluations are widely practiced in large firms. And their importance and frequency is rising continuously.[2] Most of the time, they consist in annual meetings between a superior and his subordinate. In some firms, review meetings take place at a bi-annual rate, or even more often. For example, many consultancies evaluate their consultants' performance at the end of each project completed, and this is done in addition to regular bi-annual reviews. Often, the companies commit to the conduct of periodical evaluations by anchoring them in their internal guidelines. Thus, the frequency of evaluations is generally something employees know in advance and can hence adapt to.

[2] A 2005 survey entitled *Under the microscope—staff appraisal frequency on the increase* (2005) of the IRS Employment Review states that in 2005 over 50 percent of employers carry out performance appraisal at least at an annual frequency, over a third evaluates at a six-monthly rate. In 2003, annual reviews were conducted by 70 percent and bi-annual rates were practiced by less than 20 percent of the surveyed firms.

Using economic modeling techniques, this work is a primary investigation of the role that the frequency of performance evaluations plays in incentive contracting. We approach the topic assuming a supervisor whose judgment is biased by imperfect memory. In our theoretical model, supervisors—even if they are aiming at truthful and correct evaluations—only have a bounded capacity to handle the information necessary for the evaluation process. The longer the period between the performance itself and the judgment of its value, the less likely is the supervisor to correctly remember the real value of an employee's work. Hence, this bias is likely to weaken incentives and the employer can choose a higher frequency of performance evaluations to counterbalance the loss of information on behalf of the supervisor.

Technically, the assumption that the probability of forgetting information increases over time corresponds to a depreciation of information, attaching a higher weight to more recent information than to prior. Explaining this kind of bias by a memory imperfection is only one possibility. One can also argue that a decision maker correctly remembers all information, which has been gathered in the past, and his bias consists in an over-emphasis of more recent information.[3] In this case, the bias is not generated by imperfect memory but is leads to the same biased evaluation result. The interpretation of the bias considered in this thesis will be discussed in more detail in the following chapters.

When asking professionals about their perception of performance evaluations, one will frequently hear them complain about how time-consuming the entire evaluation process is. In our analysis, this feature is captured in a cost coming along with each evaluation that is conducted. So all together, more frequent evaluations offer a benefit by enhancing the quality of information used for the determination of incentive payments but they are costly at the same time. In the following chapters, we will use this particular framework to analyze the interplay

[3]This kind of bias can, for instance, be relevant for investor behavior. A biased investor will overly weight recent investment signals.

between an employee's attitude towards risk and the frequency of evaluations, as well as the impact of heterogeneity of both, supervisors and employees.

1.2 Frequency Matters!

Before we go deeper into the analysis of the setup described beforehand, we will shortly discuss some examples which illustrate that *frequency* matters in the process of performance evaluation.

The Right to Evaluation

Periodical evaluations are prevalent for jobs where performance is mainly reflected by qualitative attributes. Middle-management positions such as business analysts and accountants, but also engineers are good examples. These employees can only be assessed by a person who has sufficient insight into the profession and the specific task they are working on. Especially in larger companies, the information gathered in the evaluations is not only of particular interest for the determination of the next bonus payment. It is also crucial for the employee when he wants to change his position in the firm or when his superior is likely to change his job. One example is a large German chemical company which attributes to her middle managers the *right* to be evaluated at a bi-annual rate. Employees can choose whether they wish to be evaluated at the maximal (bi-annual) frequency or only at an annual rate. In particular, the more frequent fluctuations are, the higher the interest on the employee side for performance information to be properly documented. Additionally, well-documented positive performance information can serve as an insurance against a negative reference when the employment relationship is terminated due to a conflict. And even more, the employee can dispose over the information fixed in the evaluation to use it as a reference when searching for a new job—inside and outside the firm!

What we really find striking in this example is that evaluations serve employees and not only employers. They can be seen as a boost to an employee's position, who at first sight is only a number in a large corporation. To firms, more evaluations are first of all costly, but the lesser evil when incentives are weakened by infrequent evaluations. To—well performing—employees (and this can be seen as a further incentive) performance evaluations are crucial, not only in determining current pay but also future career options. When employees are not used to it, the first reaction to performance evaluations generally is a rejecting one. It is associated with pressure and surveillance. But in the example above, as well as in our analysis, frequent evaluations can be advantageous rather to employees than to firms, who wish to avoid too frequent evaluations to save cost.

Evaluations and Goal Setting

Especially for sales staff, meetings defining current sales targets take place at a high rate. Weekly updates are not unusual. While these meetings are not performance evaluations themselves, they show that in some circumstances the regular update of information in response to the changes in external factors, e.g. the market situation, is crucial. If, for instance, market activity cools down due to an unexpected event and sales targets are not adjusted at a frequent rate, this creates a bias of the retained performance information and weakens incentives. Abstracting from a sales context, it is plausible to believe that external factors influencing the performance of employees will be taken into account with more precision if employees are evaluated more often. This could also be a reasonable explanation, why investment bankers are evaluated on a quarterly basis whereas lawyers' evaluations are rather rare.[4] The investment banker's exposure to external risk, i.e. his dependency on stock market performance, is much stronger than that of a lawyer.

[4]This example is taken from Lazear (1990).

Evaluations in an Educational Context

In education, the evaluation of school achievements is documented by report cards, who are themselves an aggregation of different marks assigned to former achievements. One would expect that more observations, i.e. tests, written examinations, or homework, included in the determination of a particular mark, should give a more precise description of the actual level a pupil has in the corresponding school subject and this mark will be perceived as fairer.

But not only the number of observations will be important for the quality of performance information in education. Let us consider an evident example, namely evaluations in the context of doctoral studies. The German system of doctoral studies is presently subject to an important change. Most doctoral students used to be employed at the chair of their advisor but recently more and more applicants choose to join doctoral or so-called graduate schools to obtain a doctoral degree. An important difference between the two systems of doctoral formation is the frequency with which research findings are presented to the advisor of the thesis and to other researchers. Regular research workshops taking place at the doctoral schools are one important difference of the new system compared to an employment relationship between the professor and the Ph.D. student. This is regarded as a means to accelerate and improve Ph.D. studies. Another distinguishing feature of the new practice is the information provided to the student through the regular workshops. It will play a major role in his or her future behavior whether the research paper has been approved of or not as this information directly impacts his further research strategy.[5]

[5] We will consider further details of this example in Section 3.5 and in the conclusion.

1.3 Main Results

In the simplest setup of our model (Chapter 3), i.e. when supervisors cannot retain information perfectly over time and employees are risk-neutral, we find that supervisors will not evaluate their subordinates more frequently to counterbalance forgetfulness. Although, the quality of performance information depends on evaluation frequency, this does not automatically create a trade-off between evaluation and incentive cost. Rather, the additional risk caused by the loss of information due to imperfect memory will be entirely shifted to the employee via a negative fixed compensation and incentives are strengthened by the means of high bonus payments.

The above-mentioned trade-off is created as soon as we consider workers or employees whose wealth is limited and who at the same time cannot borrow money. Such a binding wealth-constraint restricts the extent to which an employer can make use of negative fixed compensation for the sake of higher incentives. If in this case, incentives for early periods become too expensive, the employer will prefer to augment the number of evaluations. In the first instance, this reduces his incentive cost but also the surplus at stake will be lowered due to increased evaluation cost. A similar trade-off occurs when the subordinates are assumed to be risk-averse (Chapter 4). Depending on the cost of an additional evaluation, it might be worthwhile to increase evaluations as this allows the employer to lower variable compensation. By absorbing the effect of the memory imperfection, more evaluations eliminate the efficiency loss due to the risk imposed on the agent.

Extending the model with risk aversion to a duopoly setup with heterogenous *employers* (Chapter 5) does not change the decision on the evaluation frequency. Only, an employer who has a better memory capacity might have an advantage over his competitor. Since the distribution of market power does not affect the chosen evaluation frequency, efficiency is not affected by the introduction of a duopoly framework.

When *employees* are heterogenous with respect to productivity (Chapter 6), we find that the choice regarding the optimal evaluation frequency is independent of the specific productivity level if information is symmetric. In contrast to our simple model with risk-neutral employees, we find a trade-off between evaluation cost and the quality of performance information induced by the different cost function for effort, which is quadratic and not linear. Furthermore, we find that sorting the workers can be worthwhile if the firm is imperfectly informed about the employees' productivity.

1.4 Structure

Chapter 2 anchors this work in the relevant literature on contract theory, incentive provision, and performance evaluations. A simple model of an evaluation process with bounded memory and extensions to limited liability and depreciation of output will be presented in Chapter 3. Chapter 4 extends the setup by considering employees who are averse to risk. Subsequently, Chapter 5 introduces a first source of heterogeneity by dealing with supervisors who differ in their memory capacity. Chapter 6 transfers the heterogeneity from the side of the supervisor to those being supervised. Whereas in Chapter 5 heterogeneity is an observable feature of the supervisors, now asymmetric information is added. The employer knows that there are two types of employees or workers, but he cannot tell which type he actually faces. The thesis concludes in Chapter 7 with a summary of the main insights, a discussion of the results, and areas for future research.

Chapter 2

Literature

In this chapter, we discuss different strands of literature which are particularly relevant for the models developed in Chapters 3 to 6. Whereas Sections 2.1 and 2.2 give an overview of this literature, Section 2.3 highlights the main connections between the present work and the literature cited beforehand and points out further contributions playing an important role in the present work.

2.1 Contract Theory

With the introduction of a framework for the analysis of decision making under uncertainty (von Neumann and Morgenstern; 1944) the analysis of risk and risk sharing began to play a major role in the economic research of the 1950s. During the following two decades, the so-called theory of incentives and the economics of information emerged. A main driving force of the progress in both of these fields was to provide a formal framework for a theory of economic institutions such as the firm (Bolton and Dewatripont; 2005, chap. 1). So to speak, they were important contributions to the theory of the firm (for an overview see Holmstrom and Tirole; 1989) which had been initiated by the seminal article of Coase (1937). The strong link between the theory of the firm and what we call contract theory today manifests in the interpretation of the firm as a nexus of contracts (Alchian and Demsetz; 1972; Jensen and Meckling; 1976). By the 1990s, this strand of economic research had moved away from the question *why* an institution such as the firm exists towards a deeper understanding of *how* such economic institutions work, i.e. an additional theory of organizations emerged. This is one explanation why the theory of information, incentives, and economic institutions nowadays is generally referred to as contract theory (Bolton and Dewatripont; 2005, chap. 1).

2.1.1 Adverse Selection

The pioneering innovation of the theory of information was to incorporate the informational structure—the question who knows what at which point in time—in economic research. The focus of this strand of literature is to analyze how changes in this informational structure affect the outcome of economic interaction. A pioneering work in information economics is Akerlof (1970). In his paper on the lemons market, he illustrates how an informational asymmetry between buyers and sellers on a particular market could, in a fully competitive situation, lead to market failure. The general problem he deals with is referred to as *adverse selection*. One party in the market (in Akerlof's case this is the seller of a used car) is better informed than the second party (the buyer of the car), hence, the better informed party has some *private* or *hidden information* at her disposal. There are two types of economic models that deal with adverse selection, *screening* and *signaling* models (see Riley; 2001, for a survey). In both types of models, the *asymmetric information* is generally caused by an unobservable heterogeneity on one side of the market. In Akerlof's model, there are different types of sellers, some selling good, some of them bad quality. Spence (1973) proposes a model of imperfect information in the labor market where potential employees differ in their productivity. In the case of screening, the incompletely informed party first proposes a contract followed by acceptance or rejection by the party holding private information. In turn, signaling refers to a situation where the first party to act is the one with private information.

Many screening models deal with the design of contracts offered by an imperfectly informed monopolist (Vickrey; 1961; Mirrlees; 1971). As some attributes of the party he wants to contract with are unobservable, the monopolist offers a range of contracts designated to each possible type of contractual partner. If the informed party chooses the 'right' contract, i.e. the contract particularly designed for this type, the monopolist is able to screen the agents. With only two types of privately informed individuals, it is a general result that the contract for the favored type is efficient and this type can extract a positive rent from the contract. The contract for the less favored type is inefficient and he gets a utility of zero (Salanié; 1997, p. 25-26). The existence of screening equilibria in a competitive market has been studied by Rothschild and Stiglitz (1976). In their study of insurance markets, they find that an insurer will always screen individuals with different inherent risks and a screening contract only gives complete insurance to individuals with high risk.

In the case of asymmetric information and signaling, the informed party moves first. Signaling is a rational behavior if some of the informed individuals can reveal their private information using a signal. For this signal to be significant, the cost of the signaling activity must be correlated with the type of the individual, such that it is only profitable for a particular type to invest into this particular signal. Referring to a labor market context, the seminal paper by Spence (1973) analyzes under which circumstances an informed job candidate can reveal information on his productivity using his education as a signal.

2.1.2 Moral Hazard

Adverse selection refers to an *ex ante* informational asymmetry occurring before parties decide about prices and quantities traded, i.e. before they conclude a contract. In another well studied class of economic problems, asymmetric information arises only after parties have settled a contract and generally concerns information about the action taken by one party. This hidden action accruing *ex post* is referred to as *moral hazard* in the economic literature. An important issue for the study of moral hazard is delegation. Stockholders delegate the administration of firms to managers, citizens delegate governance to politicians, or employers delegate daily business to employees. The delegates, or *agents*, are generally better informed than the instructing party (*principals*) when making decisions in place of their principals. Hence, the latter cannot conclude which actions have been taken if his only observation is the outcome generated by the agent. In general, this kind of asymmetric information is unfavorable for principals as agents can exploit their additional knowledge for their own benefit, e.g. by accomplishing the delegated task with less care.

The common framework in which delegations are analyzed in economic research is the so-called *principal-agent model* (a methodological survey can be found in Laffont and Martimort; 2002). Salanié (1997) points out the following major elements of moral hazard in an agency-model. The principal's utility is affected by the agent's action, of which only an imperfect signal can be observed. Furthermore, the action the agent chooses "spontaneously" is not Pareto-optimal, reflecting that "the objectives of the parties differ" (Salanié; 1997, p. 108). To study the impact of imperfect information in this setup, the case of full information serves as a bench mark. When the principal can observe the action of the agent—and not only a stochastic outcome depending on the agent's action as well as a random variable—an efficient, that is a *first-best* contract, can be implemented. The principal conditions the remuneration of the agent on the action, which the latter will choose optimally. In this case the *surplus*, the economic gain of both parties taken together, will be maximal.

When dealing with an unobservable action of the agent, the principal can only condition the contract on the outcome, not on the action taken by the agent.

If the agent is risk-neutral, he will "sell" the firm to the agent in an optimal contract. Hence, the agent pays the principal a fix premium and becomes the residual claimant for all benefits arising.

Though it is reasonable to assume that the outcome is correlated with the agent's action, it is still subject to environmental influences and hence stochastic. Therefore, a contract based on the outcome makes the agent's earnings from the contract risky. When the agent is risk-averse, this fact creates what is referred to as the trade-off between risk and incentives in the literature (this trade-off and its empirical implications have been extensively discussed by Prendergast; 2000, 2002a,b). From the point of view of optimal insurance, a risk-neutral principal should give full insurance to the risk-averse agent by agreeing on the payment of a fixed amount in the contract. But due to imperfect information, this would induce the agent to make a choice which is not in line with the principal's objectives. So, to align incentives, the agent's remuneration must depend on outcome as well. The stronger it depends on this stochastic measure, or the stronger the incentives set by the contract are, the more costly it becomes for the principal to compensate the risk inherent to the contract. It is a general result, that contracting in the presence of moral hazard can only be *second-best efficient* when the agent is risk-averse. The magnitude of incentives—this is what determines the agent's exposure to risk—will in equilibrium be buffered according to his degree of risk aversion. Full insurance will not arise in a second-best contract.

A further aspect that is important for the efficiency of contracting in the presence of moral hazard is the wealthiness of the agent. The simple solution to sell him the firm and hence to make him residual claimant of his actions is only possible if the agent is sufficiently wealthy to buy. If he is not and if his access to capital is limited, the outcome of the contractual agreement cannot be more than second-best. Analogously to inefficiency rising with stronger risk aversion on the part of the agent, the efficiency loss also tends to be aggravated the lower the agent's wealth is (see e.g. Bolton and Dewatripont; 2005, p. 168).

James A. Mirrlees authored several early contributions to optimal contract design with moral hazard (Mirrlees; 1974, 1975, 1976). Further pioneering papers in a principal-agent setup are Ross (1973), Harris and Raviv (1979), Holmstrom (1979), Shavell (1979), and Grossman and Hart (1983).[1] The papers cited beforehand do not assign a particular form to the contract offered by the principal. In practice, it is quite common that the principal pays the agent a fixed premium plus a bonus depending on the outcome. Holmstrom and Milgrom (1987) have made an attempt to rationalize such linear contracts, though they find that very strong assumptions must hold for them to be optimal. Still, the wide use of linear contracts and their convenience for economic modeling make them a popular field of study within contract theory. A good example is the work on piece-rate pay which we will discuss in Section 2.2.2.

2.1.3 Economic Contracts: Definition and Distinctions

In economic theory, what is referred to as a contract is a wide range of agreements. Schweizer (1999) describes these agreements as a much broader concept than only the contract in a legal sense. Rather, economic contracts include all institutional precautions defining, influencing, and coordinating the strategic behavior of interacting parties (Schweizer; 1999, p. 5).

An essential distinction within economic contracts is the one of *complete* and *incomplete contracts*. For a contract to be incomplete, Tirole (1999, pp. 743) identifies three potential conditions: unforeseen contingencies, the cost of writing contracts, and the cost of enforcing contracts. When one or several of these conditions apply to a contract, it is said to be incomplete, otherwise it is complete. Some major contributions especially discussing the role of property rights when contracts are incomplete are Grossman and Hart (1986), Hart and Moore (1990),

[1] An important concept introduced in the papers by Shavell and Holmstrom is the *informativeness principle* which stipulates that a contract should optimally include all available information on the agent's action.

and Hart (1995). In particular, they analyze under which circumstances ownership is an appropriate way to overcome opportunistic behavior on behalf of the contracting parties caused by contractual incompleteness.

The verifiability of the information the contract refers to plays an important role for the economic analysis of contracts. First, the more difficult it is to verify this information, the larger the cost of enforcing the contract. Second, it is also plausible to think of situations where it becomes so difficult to verify the information that the respective cost is prohibitive. Basically, this results in the information being *unverifiable*. Such a contract, containing unverifiable information, is referred to as an *implicit contract*. In contrast, we speak of an *explicit contract* if the information used in the contract can be verified by an institution which has the power to enforce the contract in case of an infringement or a disagreement on the terms of the contract. Most contracts the economic literature deals with are explicit. This is obvious as the incentives to stick to an explicit contract are guaranteed by the enforcing institution. In contrast, in an implicit agreement the contracting parties can decide on common terms, but they know that this accord will not be protected by the threat of a legal dispute that puts through the terms initially agreed. Hence, an incomplete contract seems to be a fragile construct and one would expect not to observe individuals to agree on using such contracts as it is not credible that they will stick to them. But this is not always true. Especially in labor contracts, implicit agreements between employers and employees are quite common. Even though explicit contracts dominate the economic discussion, the analysis of implicit contracts has attracted more attention since its emergence at the end of the 1970s.

The term 'implicit contract' goes back to Azariadis (1975); synonymously, the term *relational contract* is used. A distinctive characteristic of implicit contracts is that parties are normally not willing to stick to them in a one-shot interaction as the terms of the contract are not enforceable. This might change when the duration of a relationship is open-ended. If the economic surplus generated by a continued compliance of all parties to the contract is sufficiently high, the parties will stick to it, even if there is no institution guaranteeing enforcement. In this case, the termination of the contractual relationship, or a contractual pause for several periods of time whenever an infringement of the contract occurs, is a sufficient threat of punishment motivating the parties to adhere to the implicit contract. The value of the surplus from future contracting being sufficiently high, the implicit contract is *self-enforcing*, i.e. no third party is needed to assure compliance. So, in this situation, the parties are able to build up a *reputation* of compliance to the implicit contract and, hence, they can *trust* each other. The game-theoretic concepts allowing to prove the existence of such self-enforcing implicit contracts are the so-called Folk Theorems (e.g. Friedman; 1971; Rubinstein; 1979).

Klein and Leffler (1981) particularly discuss implicit contracts in the context of product quality on a goods market. If market prize sufficiently exceeds marginal cost, an implicit agreement concerning the level of quality will be self-enforcing, even when the quality of goods is assumed to be unverifiable ex post. The papers by Bull (1987) and MacLeod and Malcomson (1989) are important contributions to the analysis of implicit contracts in labor markets. Bull shows that not market reputation but an employer's intra-firm reputation is the key for the existence of self-enforcing termination contracts assuring an efficient level of effort. In a certain sense, MacLeod and Malcomson approach the subject in a more general manner showing that implicit contracts can consist of various agreements, e.g. of piece-rate or bonus contracts, and are not restricted to termination contracts, where a high wage is payed as long as employees exert sufficient effort and otherwise the contract is terminated. Furthermore, a rigorous study of implicit contracting under moral hazard and asymmetric information can be found in Levin (2003).

2.2 Contract Theory and Incentive Provision

An important application for contract theory is the design of incentives in firms (see Prendergast; 1999, for a comprehensive overview on this subject). Monetary employee incentives are generally determined by the choice of compensation contracts which can for example comprise a fixed compensation, pay for performance, or both.

2.2.1 Fixed Compensation

A quite common compensation scheme is to pay the employee a fixed salary, independent of the output being produced. The seminal paper by Shapiro and Stiglitz (1984) shows why this type of contract, even if there is no direct link between compensation and output, provides incentives for workers to exert some positive effort level. The rational is the following. In this model employees are being paid above their marginal productivity. Their effort level cannot be observed but with a positive probability monitoring will discover when an employee has shirked, i.e. when he has fallen short of a minimal effort level. He will be fired in case of insufficient effort exertion. The more the wage exceeds the agent's marginal productivity, being fired becomes a stronger punishment for him. He will hence have an incentive not to undercut the minimal effort level to avoid being fired. Though, this kind of fixed-wage contract, referred to as *efficiency wage* in the literature, does not directly link pay to performance, it clearly sets some incentives by the threat of dismissal.

Fixed salaries are generally paid in return to a given number of hours worked. Thus, a fixed-wage compensation scheme compensates employees according to their input measured in hours. The salary also happens to vary with the hours worked, e.g. if overtime work is paid or working hours are flexible to some extent. Still, this can be considered as a fixed-wage compensation scheme in a narrow sense as long as the hourly wage is constant. What is generally referred to as variable compensation varies with output not input. We will discuss the effects of compensation being linked to output in the following section.

2.2.2 Pay for Performance

Whereas a salary compensates a worker for the provided input, compensation solely based on output is called a piece rate. The introduction of an efficient piece-rate incentive scheme usually has two major effects. It strengthens the incentives of the employees remaining with the firm, but it also induces a selection effect. Less productive employees might choose to quit the current employer in search of a firm paying a straight salary, while more productive employees on the labor market are attracted as they expect to be paid more via piece rates than under a fixed wage schedule. Lazear (2000) quantifies both effects using data from a large auto glass company switching from hourly wages to piece rates.

A theoretical analysis of the impact of piece rates on the incentives of the existing workforce of a firm, their effect on employee retention and on hiring options can be found in Lazear (1986). The paper identifies several factors favoring compensation by piece rates over paying a straight salary. When a firm's workforce is rather heterogenous with respect to ability a differentiation of compensation according to the generated output will provide incentives at a lower cost than a straight salary. In other words, when a worker's reservation wage sufficiently exceeds the average output at his current firm, he can only be efficiently retained by paying him a piece rate. Of course, this holds only if paying a salary means paying the same amount to all workers doing the same job for the same number of hours.

According to Lazear, two further determinants for the choice of the compensation scheme are the cost of measuring output and the accuracy of performance measurement. If it is difficult to provide a significant measure of individual output at a reasonable price, paying a piece rate will either provide low or wrong incentives, or it comes at a high cost. So, the option to pay a straight salary should be considered. In the following section we will further discuss characteristics of performance measures and in how far they impact efficient incentive design when used in performance evaluations.

2.2.3 Performance Measures and Evaluations

Paying for performance needs adequate performance information. A means to form the link between actual performance and the corresponding compensation is to evaluate the employee. Especially in larger firms, performance appraisals or evaluations are practiced at a regular rate. These evaluations have a number of different functions (Backes-Gellner et al.; 2001). The evaluations can simply serve to determine the competencies of the workforce, e.g. by testing the intelligence quotient. Subsequently, such tests can be used to support decisions of promotion or dismissal. Evaluations also allow to give feedback to employees so as to enable them to adapt future behavior correspondingly. Last but not least, evaluations also serve as a motivation device by determining monetary incentives. They can do so in two ways, which are rather similar when contracts are self-enforcing. On the one hand, using the information revealed in the evaluation, one can adapt current pay to past performance, i.e. give a wage raise for good performance or cut wages if poor performance is stated. On the other hand, the amount of a variable compensation component can be determined by considering which values the underlying performance measures have taken in the evaluation.

Performance evaluation and performance measurement are highly related issues. The measures themselves are useless if they are not linked to decisions determined in the evaluation. So the evaluation itself is a measurement in a certain sense. Backes-Gellner et al. demonstrate this fact with the following example (Backes-Gellner et al.; 2001, p. 540): if workers are paid according to individual output, the latter must be measured beforehand. As it determines the worker's wage, the measurement itself is an evaluation of the particular employee's performance. At a regular rate, the worker is evaluated using the retained number of items produced by him and this results in a contingent payment. But one could also base the payment of the worker on overall production divided by the number of employees participating in the production process. This would avoid the measurement of individual output, but linking pay to aggregate output surely weakens the individual worker's incentives. So even though the measures for performance are available, paying piece rates will only motivate employees to perform if their individual performance is assessed in an evaluation. Hence, there cannot be performance pay without performance evaluation.

In the accounting literature there is a strong focus on performance measurement, discussing how different characteristics of performance measures used in evaluations affect incentives. This literature distinguishes between *objective* and *subjective* performance measures.

Objective Performance Measures

As we have already mentioned in Section 2.1.3, the verifiability of information contained in a contract matters. An objective performance measure can be verified unambiguously by a court when it comes to a legal dispute about the compensation to be paid to an employee. Hence, such a measure allows to base an explicit contract on it. Good examples for objective performance measures are the number of newly canvassed customers, earnings before interest and tax, or quarterly sales.

Objectivity of a performance measure does not ensure that the measure in question is a good measure of an employee's actual performance. Technically spoken, the higher the correlation between the value he generates and the performance measure the higher the quality of the performance measure. If a performance measure is not only affected by the actions and decisions of the employee whose pay is tied to the measure, but also varies with factors the particular employee cannot influence, this exposes him to a risk he must be compensated for. This is the fundamental inefficiency in principal-agency relations described by Holmstrom (1979) and others (Section 2.1.2 addresses this issue in a more detailed way).

Another important characteristic of a performance measure is whether it can be manipulated, i.e. whether the employee can boost the measure by actions that do not necessarily create additional value for the firm. This can be illustrated by considering the use of profit as a performance measure. If the manager of a production plant is paid according to profit, he has a large number of means to augment this measure. While the firm is interested in long-term profitability, the manager might choose to omit important investments at the production site in order to be paid a higher bonus in the short term. Usually, the firm will have a longer time horizon than its employees, and hence, sustainable profitability is not ensured by linking pay to current profitability. A theoretical analysis of incentive contracts that are tied to an inefficient, i.e. a distorted performance measure, is provided in Baker (1992). It turns out that efficiency depends on the statistical relationship between the underlying objective of the employer and the performance measure used in the contract.

This is also true in practice. What you pay for is what you get! Employers introducing a variable compensation contingent on a distorted performance measure risk making harsh experiences. High values of the performance measure might be realized to the detriment of the firm or of customers. A well-known example for this effect is the introduction of piece rates at Sears' auto repair centers, which resulted in a substantial reduction of the firm's value (Patterson; 1992). A short description of the case can be found in Baker (2000, p. 416.).

Subjective Performance Measures

In order to avoid counterproductive or suboptimal behavior objective measures are frequently combined with or even replaced by subjective measures. Calling in a supervisor to provide his personal judgment allows to filter out some risk imposed by objective measures, it reduces distortion, and constitutes a means to investigate whether there has been "gaming" of objective performance measures. The judgment of the superior being unverifiable, an incentive contract based on subjective performance measures is implicit, entailing the problems we have already discussed with respect to implicit contracts in general in Section 2.1.3. Therefore, the existence of a sufficient amount of trust between the contracting parties is crucial for the functioning of incentive schemes based on subjective measures (see e.g. Harbring; 2007, for an experimental study on the role of trust and cooperation in organizations). Recent empirical evidence in Gibbs et al. (2004) supports this by showing that the positive impact of subjective evaluations on productivity, profit, and employee satisfaction increase in tenure of the supervised.

It is often argued that only weak monetary incentives are tied to the outcome of subjective evaluations. But one should not forget that variable compensation at the end of an evaluation period can be chosen equivalently to the expected discounted gain from promotion due to a good performance in the past evaluation period, the timing of the payment being irrelevant. So, monetary incentives are still at work if promotion decisions coming along with a wage raise are influenced by evaluation outcomes.

Objective and Subjective Performance Measures

Baker et al. (1994) have formally studied the interplay between the use of subjective and objective performance measures and find that the simultaneous use of both, implicit and explicit incentives, can be optimal. Related to this work is a paper by Schmidt and Schnitzer (1995) who extend MacLeod and Malcomson (1989) to a setup with implicit and explicit contracts at work. In line with Baker et al. (1994), they show that an objective measure which is close to perfect may destroy the possibility to contract on a perfect subjective measure. So, to be able to use subjective measures in performance evaluations, the objective measures at hand must be sufficiently imperfect. Further work focussing on the interaction of implicit and explicit incentives in a principal-agent model is Pearce and Stacchetti (1998). When some actions of the agent are non-verifiable, rewards based on an implicit contract can smooth the agent's consumption and, hence, alleviate the risk imposed on the agent.

Budde (2006) can show in an extension of Baker et al. (1994) that implicit and explicit contracts are in fact complementary, no matter what is the principal's fallback position. The balanced scorecard, a concept going back to Kaplan and Norton (1992), underpins the practical relevance of this result. Nowadays, it has become a popular management tool that implements the combination of objective as well as subjective performance measures to support operative control and incentive provision in firms.[2]

Subjective Performance Evaluations and Bias

Beyond issues concerning the reputation of an employer, the use of subjective performance measures comes along with some more inconveniences, namely the discretionary character of subjective evaluation methods and the bias of measurement they might entail. Prendergast and Topel (1993) discuss different types of bias that can influence the subjective judgment of a supervisor. For example, there is evidence that supervisors give higher ratings to subordinates of their own

[2]It has to be mentioned, that the use of scorecards for incentive provision is also subject to criticism. Lipe and Salterio (2000) as well as Ittner et al. (2003) find that superiors tend to attach too much weight to objective measures within a scorecard. Furthermore, Moers (2005) gives empirical evidence from a one-firm sample that indicates bias in evaluations being reinforced by performance measure diversity and subjectivity.

race. Also, personal affection of a supervisor might lead to better ratings of favorites (this situation has been modeled in Prendergast and Topel; 1996). The mere existence of favoritism naturally leads to influence activities designated to win the favor of the supervisor, creating incentives to invest effort in activities that—at least most of the time—do not serve the objective of the organization.

A further important contribution to the study of bias in subjective performance evaluations is MacLeod (2003). In this paper, both, the supervisor and the subordinate, privately observe a subjective performance measure. In the evaluation process, they must agree on a single evaluation result. This becomes more and more difficult as the correlation of the two privately observed performance measures declines. Disagreement is costly, as the subordinate is assumed to punish the supervisor in case he has been treated unfairly, i.e. if his subjective measure indicates a better rating than the measure observed by the supervisor.

2.3 This Work and the Literature

In the two preceding sections, we have viewed a large body of literature in a quite general way. Furthermore, some particular strands of literature should be mentioned, though an extensive overview would be too comprehensive to be presented here. The present section shortly relates this thesis to contract theory in general and to the literature on incentive provision in particular, both of which we have already addressed in the above sections. In addition, we will take a glance at the link between this work and the literature on bounded memory (Section 2.3.1). Finally, we will refer to literature studying frequency issues from an economist's point of view in Section 2.3.2.

Contract Theory

The models we will present in the following chapters are supposed to reproduce interaction in situations of delegation. Therefore, all of them will be set up in a principal-agent framework. Though, most of the time the agent's effort is assumed to be observable, the information regarding effort exertion is potentially subject to a distortion due to the memory limitations of the principal. Whereas in a classical principal-agent setup with moral hazard, the risk imposed on the agent stems from an environmental shock, in our setup the risk depends on the way evaluations are conducted. If the principal takes the cost to evaluate at the end of every period, he has perfect information on the agent's performance and can adjust incentives efficiently. With fewer evaluations, the agent must be paid a risk premium but evaluation cost is lowered.

As we restrict the analysis to linear compensation contracts (e.g. Holmstrom and Milgrom; 1987), we do not provide an exhaustive study of optimal contracting with different evaluation frequencies. Assuming that they are already in use, we rather investigate in how far the design of linear contracts is affected by variations in the frequency of evaluations.

Starting with a model without hidden action and risk aversion (Chapter 3), we will gradually introduce more complexity in Chapter 4. Though, heterogeneity is already dealt with in Chapter 5, problems with respect to hidden information and screening in particular will not be considered until Chapter 6, where the principal cannot distinguish between agents of different productivity.

Contract Theory and Incentive Provision

Of course, the literature on the use of performance evaluations is particularly important for this research. By analyzing the advantages and drawbacks of different types of performance measures, we can better understand for which type of occupation evaluations based on subjective performance measures are susceptible. Also, the impact of measurement accuracy can be transferred to our setup as figuratively, the frequency of evaluations enhances measurement accuracy by balancing the effect of bounded memory.

In Section 2.2, we have referred to work that includes the existence of a bias in the analysis of performance evaluations. Both works, Prendergast and Topel (1996) and MacLeod (2003), derive the equilibrium outcome without integrating reputation into the models. The setup we propose is relevant only for sufficiently long time periods, which allow the creation of trust between the supervisor and the employee being evaluated.[3] The study by Gibbs et al. (2004) outlining the importance of tenure for the effect of subjective performance evaluations is important evidence supporting our approach.

2.3.1 Evaluations and Memory

Since the path-breaking work of Herbert Simon on bounded rationality (e.g. Simon; 1955), a large strand of economic literature makes the attempt to explain the gap between behavior labeled as "rational" by economists and the behavior that is observed. An important contribution to this development has been made by the integration of concepts from neighboring social sciences in economic research (e.g. Kahneman; 2003). The study of incentive provision is a good example. Whereas economists have focussed on efficiency under rational behavior, other social sciences have analyzed behavior in response to incentives, but also studied in how far environmental conditions as well as characteristics of an individual conducting evaluations influence the evaluation outcome (e.g. D'Andrade; 1974, searches for systematic memory bias in the assessment of behavior).

In the present work, we analyze evaluations carried out by a supervisor who is assumed to have imperfect memory. In economic modeling, there are different

[3] Conditions for self-enforcing contracts are considered in Chapter 4.

ways to deal with memory imperfections. E.g. Piccione and Rubinstein (1997) consider implications of imperfect recall in extensive decision problems with a single player. In their setup, the forgetful player knows, that he might have forgotten an action he has taken in the past, but he is aware of all possible histories he might face. By contrast, Mullainathan (2002) provides a theoretical model of bounded memory where the decision maker can be considered naive, i.e. he acts as if he had not forgotten. Though using different approaches, both papers emphasize situations where the loss of information due to imperfect memory results in inefficiency.[4]

The situation we analyze in this thesis is quite close to work by Sarafidis (2007). He characterizes the behavior of an individual that is rewarded on the basis of an assessment of his past performance on a fixed date. The assessment is provided by a supervisor who must rely on his imperfect memory to evaluate past informative events. The supervisor in Sarafidis' model is somewhat a receptor who processes information according to a sophisticated model of memory, but without any possibility to act. In contrast, the models presented in the following chapters allow the principal to decide how often he wishes evaluations to take place, so he has a means to cover up for his own bounded memory.[5] In Sarafidis, memory is governed by two effects: information which is received more recently and information which is received more frequently is stored with a higher probability. This thesis only considers the former effect which consists in more recent information being available from memory with a higher probability.[6] In the long run, this would be in line with the result presented in a paper on information processing with bounded memory by Wilson (2004). It derives that for longer

[4]The opposite case is stressed in Frey (2005) where the focus is rather on information individuals would like to forget, but they are unable to do so.

[5]Also in Mitusch (2006), the player is aware of his forgetfulness and can strategically rehearse information he judges to be particularly important for future decisions.

[6]Of course, this automatically puts limitations on the interpretation of the results as being introduced by bounded memory. On the other hand, this simplistic approach has the advantage that it easily allows us to view the results in the light of alternative interpretations.

sequences of information a decision maker with bounded memory will put too much weight on most recently received information.

2.3.2 Evaluation Frequency

Frequency issues are not prevalent in economic research due to the fact that it is hard to give tangible answers to questions concerning frequency. In search of an optimal frequency, the results generated by microeconomic models strongly depend on the particular circumstances. An example for the study of optimal frequency in public economics research is Akemann and Kanczuk (2003) who analyze the effects of term lengths on welfare to gain insights on the optimal timing of elections. Also, frequency is an important topic with respect to the analysis of optimal monitoring policies (e.g. Ichino and Muehlheusser; 2004). Often, the frequency of monitoring is modeled via the probability that an individual's actions are controlled.

Mitusch (2006) investigates under which circumstances a supervising manager will invest in additional costly monitoring if he has to justify a bad rating of his subordinate in a performance evaluation with evidence. Whereas evidence for poor performance discovered by monitoring is verifiable, the manger cannot commit to monitoring as such, as it is unverifiable when no negative evidence is found. Hence, a contract that engages the manager to monitoring is not enforceable and the degree of supervision will depend on the respective incentives of the supervisor to do so. The optimal incentive contract depends on the monitoring cost, but also on the ability of the subordinate to engage in information dissipation to convince the manager that costly monitoring is not necessary.

Whereas it might be difficult to commit to monitoring, we believe that evaluations do not take place according to a probabilistic process or the discretion of a supervisor. Rather, the frequency of evaluations is associated with deterministic events as, especially for large firms, evaluations are part of internal guidelines.

To our knowledge, Lazear (1990) is the only work formally addressing this issue. Just as in the different models presented in this thesis, this is done by varying the frequency of performance evaluations in a simple two-period setup. Though, both papers put the question why the frequency of performance evaluations differs between professions or firms, and though, both explain a large part of the differences by an informational effect of evaluations, the focus of Lazear's rather unknown early work is quite different to the present thesis. Whereas evaluation frequency in our setup impacts the quality of the performance measure used in the evaluation, in Lazear (1990) more frequent evaluations reveal information on the employee's productivity potential at the current firm, and hence, allow the employee to eventually switch jobs at an earlier state if doing so is profitable. As in the models presented in this thesis, in Lazear's setup, the efficiency gains of more frequent evaluations are traded off against the cost caused by this time-consuming procedure.

Chapter 3

Evaluations and Bounded Memory

3.1 Introduction

This chapter investigates the role of the frequency of performance evaluations in a simple incentive contract based on subjective performance evaluations. In our model supervisors, even if aiming at truthful and correct evaluations, only have a bounded capacity to handle the information necessary for the evaluation process. This is plausible as the main task for superiors of middle-management and engineering staff is rarely to supervise and evaluate employees, but this is done additionally to their day-to-day work. When, furthermore, the observed information on the output first needs to be analyzed to be useful for an evaluation, we expect to find a bias in performance evaluations due to imperfect memory.

The supervisor in Sarafidis (2007) is in a sense a receptor who processes information according to a sophisticated model of memory, but without any possibility to act. In the following model, however, he can decide how often he wishes evaluations to take place, so he has a means to work against his own bounded memory.

We model only an effect which consists in more recent information being available from memory with a higher probability. We will take the feasibility of implicit contracts based on subjective performance as given and rather explore the specific impact of bounded memory, i.e. a systematic bias in evaluations due to a *recency effect*[1]. Like Akemann and Kanczuk (2003) and Lazear (1990), the following model subdivides the entire time span in two periods to allow for comparison of different choices concerning evaluation frequency.

In this two-period setting, a forgetful employer must decide how often he should evaluate an employee who works for him. At a date which has been fixed in advance, the employer attaches a subjective performance measure to the information that he has gathered over the preceding evaluation period. We argue that, depending on the length of this period, the quality of the information he has stored varies due to bounded memory, so he probably has a rather precise idea on how well his subordinate performed in the recent past, but events that are more remote are also more likely to be forgotten. More frequent evaluations come at a higher cost, but at the same time they also lead to more precise information on which the principal can base the bonus payment for the agent. Given the contract designed by the principal, the agent decides in each period whether to exert high or low effort.

To gain additional insights, we also investigate the impact of a wealth-constrained employee. This implies that the agent receives a fixed compensation component that is bounded from below—for instance due to a minimum wage—and, therefore, can be higher than it would be optimal for the principal. The design of the underlying principal-agent model is very close to Laffont and Martimort (2002) who also introduce a wealth constraint in a discrete model. They refer to the constraint as *limited liability* and interpret this restriction as a first step towards risk aversion.

[1]This is the component of imperfect memory in Sarafidis (2007) which implies that more recent information is stored with a higher probability.

In the unrestricted model, we find that the principal never decides to evaluate the agent in each period. His forgetfulness implies a higher variable payment to the agent compared to perfect memory but he reincorporates this loss via a negative fixed payment and, therefore, he can extract the entire surplus from the employment relationship. For a wealth-constrained agent, the fixed compensation component must be larger than the one optimally chosen by the principal. We show that in this case, the formerly derived results will be affected in favor of more frequent evaluations. Unlike before, part of the surplus possibly goes to the agent.

In Section 3.2, we describe the setup of the basic model and how we will solve it in the following sections. Section 3.3.1 derives the optimal choices of the principal and the agent for the case of a principal who evaluates the agent's performance at the end of the last period. We repeat this analysis for the case of two evaluations in Section 3.3.2. In Section 3.3.3 we investigate how the choice of evaluation frequency changes if the agent is assumed to be wealth-constrained. Section 3.4 sums up the results of the previous sections. Section 3.5 is a digression and deals with the impact of a modified profit function, for which we provide a distinct intuition. Section 3.6 concludes with the main insights on evaluation frequency we can extract from the different model versions when an agent faces a forgetful principal.

3.2 The Model Setup

In this section, we develop a model that shall capture a relatively long time span, e.g. one year, in an ongoing employment relationship between a principal and an agent. The principal, who faces imperfect memory, must decide on the frequency of performance evaluations over the entire time span when he designs a contract. In the model, this time span is subdivided into two periods to allow a comparison of different choices concerning evaluation frequency. Even though we consider a two-period setting, we exclude discounting from the analysis to focus on the effect of imperfect memory on evaluation practices.

The principal, who disposes of a production technology, has employed an agent to be able to realize a surplus. But employment alone is not sufficient for the surplus to materialize, additionally the employed agent needs to exert a certain level of effort. Therefore, he has to be compensated for the disutility of the effort necessary for the creation of the surplus. This is done by the use of a linear contract, where a bonus depending on the agent's performance is paid in addition to a fixed base wage. Besides compensation, the principal possesses another instrument he can use to design an optimal contract in our setup: by investing in more frequent evaluations he can raise the quality of the available performance information and, thus, reduce variable compensation. Or to put it another way, he can decide to evaluate the agent less frequently but the additional risk that emerges from worse information about the agent's effort provision has to be outweighed by higher incentive payments.

The principal can offer a contract that determines the payments the agent receives for the output he generates on behalf of his employer. The contract is valid for the entire time span and in each period the agent decides how hard he works (high or low effort). It is common knowledge between the agent and the principal that if the agent exerts high effort (e_H), he produces high output (π_H) with certainty, that is $e_H \Leftrightarrow \pi_H$. Analogously we have $e_L \Leftrightarrow \pi_L$ for low effort (e_L) and low output (π_L). We define $\pi_H, e_H > 0$ and for simplicity, we set low effort and low output to zero, $e_L = \pi_L = 0$. To express these properties in the model, we choose the linear function $\pi_i = f \cdot e_i$, with $i = L, H$, to describe the relationship between the agent's effort e_i and the output π_i. The parameter f stands for the productivity of the agent, which is assumed to be larger than one. As the output the agent generates with low effort is zero, the principal always wants the agent to work hard. An incentive compatible contract therefore implies the choice of e_H on part of the agent in both periods.

Both contracting parties are risk-neutral and compensation is paid via a fixed component α as well as a variable component depending on a bonus factor β. Thus, we analyze linear contracts only. Compensation based on β is tied to the agent's performance, which is the output he generates during the two periods.[2] We assume that the principal is able to observe the unverifiable output in each period, but he cannot immediately translate it into a quantifiable measure that enters the bonus function. First, the employee's achievements need to be reviewed, quantified, and made explicit through a performance evaluation. So a costly thinking and classifying process must be undertaken to translate the observed performance into explicit quantitative terms that reflect the impact of performance on the principal's profit. This process is what we call the evaluation of the agent's performance. An important assumption is that the employer cannot remember information perfectly over time as long as the information he has observed has not been made explicit through an evaluation. So a principal who decides to save the cost of evaluating his agent immediately, might forget the performance he observed. In the model, this is represented by the principal remembering the observed output only with some probability ρ in the following period, if not both, production of output *and* an evaluation, have occurred previously.

When there are gains from contracting, the principal must decide how often to evaluate the agent: each period (twice) or at the end of the last period (once). Each evaluation comes at a fixed cost c_E. So on the one hand, more frequent evaluations are more expensive, but on the other hand, they lead to more precise information on which the principal can base the bonus payment for the agent.

[2]Note that the terms *performance*, *effort*, and *output* can be used as synonyms as long as the production process is deterministic.

If one of the parties refuses to accept the contract, there will be no employment relationship.

The chronology of events is the following. First, the principal designs a contract which contains the number of times the agent is evaluated, the amount of a fixed wage component and the bonus depending on the evaluated performance. The agent accepts the contract if his expected utility from the incentive contract is at least as high as his outside option, otherwise he rejects the offer. Once accepted, the contract lasts for two periods, and in each period the agent decides whether to work hard. The principal observes and receives the output the agent produces and evaluates him truthfully on the basis of what he observed. But due to imperfect memory, the principal remembers the observed output only with some probability if there has been no evaluation in the same period. To summarize, we give a brief overview of the events:

- *Period* $\tau = 0$: if there is a positive surplus, the principal designs a work contract which contains the number of times the agent is evaluated, a fixed wage component α, and a factor β that, when multiplied with the agent's evaluated performance, determines variable compensation; the agent accepts or rejects the contract

- *Period* $\tau = 1$: the agent exerts effort $e_1 \in \{e_L, e_H\}, e_L = 0, e_H > 0$ to produce the first-period output π_1; the fixed wage α is paid out

- *Period* $\tau = 2$: the agent exerts effort $e_2 \in \{e_L, e_H\}, e_L = 0, e_H > 0$ to produce the output π_2; the bonus payment is calculated according to Alternative 1 or 2 and paid out to the agent

- *Alternative 1: One Evaluation*

  - the fixed wage is denoted by α^o

- when it comes to the evaluation of the agent's work at the end of the second period the principal remembers the observed output in period one with probability $\rho \in]0,1[$
- a fixed evaluation cost c_E arises at the end of the second period
- the agent is paid his bonus $\beta^o \cdot (\rho\pi_1 + \pi_2)$ at the end of the second period

• *Alternative 2: Two Evaluations*

- the fixed wage is denoted by α^t
- the principal evaluates at the end of each period and therefore correctly remembers the agent's output he observed over the two periods
- his total evaluation cost increases to $2 \cdot c_E$
- the agent receives a bonus $\beta^t \cdot (\pi_1 + \pi_2)$ for his output over two periods

With this information at hand, we can write down the expected profit function of the principal in the case of one evaluation (superscript: o) and two evaluations (superscript: t):

$$E[\Pi^o(\pi_1, \pi_2)] = \pi_1 + \pi_2 - \alpha^o - \beta^o(\rho\pi_1 + \pi_2) - c_E,$$
$$E[\Pi^t(\pi_1, \pi_2)] = \pi_1 + \pi_2 - \alpha^t - \beta^t(\pi_1 + \pi_2) - 2c_E.$$

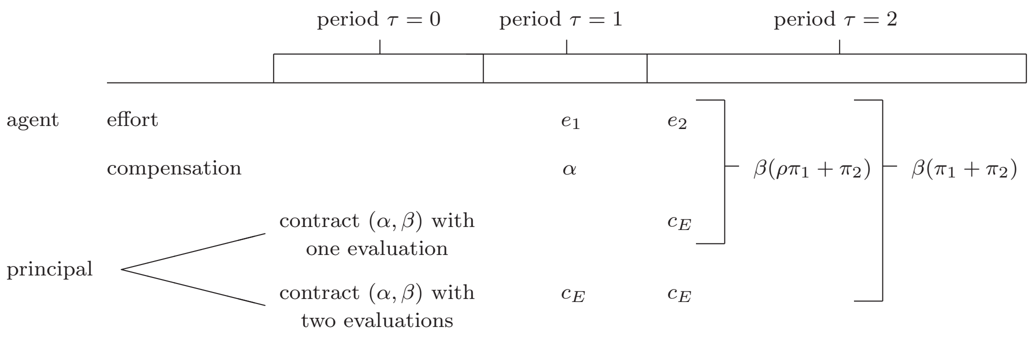

Figure 3.1: The Chronological Order of the Game

The expected utility of the agent takes into account compensation as well as the disutility from the provision of effort:

$$E[U^o(e_1, e_2)] = \alpha^o + \beta^o(\rho\pi_1 + \pi_2) - (e_1 + e_2),$$
$$E[U^t(e_1, e_2)] = \alpha^t + \beta^t(\pi_1 + \pi_2) - (e_1 + e_2).$$

The principal and the agent receive zero utility if they choose the outside option, i.e. $\bar{U} = 0$.

3.3 Analysis of the Model

We solve two versions of the model—a version with one evaluation and a version with two evaluations—for subgame perfect Nash Equilibria via backward induction. In the course of the backward induction of each model version, we derive under which circumstances the agent chooses to work hard for given compensation parameters α and β. Moreover, we determine the parameter combinations in which he will be willing to work at all. If the principal's expected profit is positive, he will choose those parameters (α and β) which yield the highest return. Otherwise he will refrain from offering a contract. After deriving the parameters for the model with one and with two evaluations, we will consider the principal's decision on the evaluation frequency that he fixes in the contract in period $\tau = 0$. We repeat the analysis for a wealth-constrained agent—this implies that α must be larger than some lower bound—and examine how such an assumption changes the preceding results.

3.3.1 One Evaluation

Given the fact that the principal evaluates performance only at the end of the second period and, therefore, has imperfect recall (represented by the parameter ρ), we now investigate for which payment schemes the agent is willing to exert high effort. We then derive the corresponding profit for the principal and deduce his optimal choice of the contract parameters for one evaluation. Independent of the actual effort levels he chooses, the agent's expected utility over the two periods can be expressed by:

$$E[U^o(e_1, e_2)] = \alpha^o + \beta^o(\rho\pi_1 + \pi_2) - (e_1 + e_2)\,. \tag{3.1}$$

And the principal's expected profit is:

$$E[\Pi^o(e_1, e_2)] = \pi_1 + \pi_2 - \alpha^o - \beta^o(\rho\pi_1 + \pi_2) - c_E\,. \tag{3.2}$$

Now, we calculate the critical parameter values for which high effort is chosen. The incentive constraints for the corresponding first-period effort level ($IC1$) and for the second-period effort level ($IC2$) must be fulfilled. Furthermore, the participation constraint (PC) must hold. So the agent will choose high effort in both periods, i.e. $e_1 = e_2 = e_H$, if the following conditions are satisfied:

$$(IC1): \quad E[U^o(e_H, e_2)] \geq E[U^o(e_L, e_2)] \Leftrightarrow \beta^o \geq \tfrac{1}{f\rho}\,, \tag{3.3}$$

$$(IC2): \quad E[U^o(e_1, e_H)] \geq E[U^o(e_1, e_L)] \Leftrightarrow \beta^o \geq \tfrac{1}{f}\,, \tag{3.4}$$

$$(PC): \quad E[U^o(e_H)] \geq E[U(0)] = 0 \Leftrightarrow \alpha^o \geq [2 - \beta^o f(1+\rho)]e_H\,. \tag{3.5}$$

As $\rho \in]0, 1[$ holds by definition, we can infer that $\frac{1}{f\rho} > \frac{1}{f}$ holds as well. Hence, only ($IC1$) in equation (3.3) is binding. A profit maximizing principal will implement the desired effort level at the lowest cost possible. If we set $\beta^o = \frac{1}{f\rho}$, we can plug this into equation (3.5) to calculate the lowest α^o in an incentive compatible contract that implements high effort:

$$\alpha^o = \left[2 - \frac{1}{f\rho} f(1+\rho)\right] e_H = -\frac{1-\rho}{\rho} e_H\,. \tag{3.6}$$

Since the principal is expected to forget part of the agent's first-period performance, it is more costly to require high effort in the first period than motivating the agent in the second period. The optimal fixed component α^o in the contract is negative, so the agent has to pay the principal to be allowed to work for him. We will analyze restrictions to this possibility using a more general framework in Section 3.3.3.

Given these contract parameters, we can calculate the principal's profit when he proposes an incentive compatible contract:

$$E[\Pi^o(e_H)] = \quad 2(f-1)e_H - c_E \, . \tag{3.7}$$

We have just derived the profit of the principal and an optimal pair of contract parameters. Additionally, offering a contract has to be at least as good as refraining from it, i.e. choosing the outside option:

$$E[\Pi^o(e_H)] \geq 0 \Leftrightarrow f \geq 1 + \frac{c_E}{2e_H} := f^o \, . \tag{3.8}$$

Condition (3.8) expresses the extend to which the agent's productivity f must be larger than one for the employment relationship to be profitable.

3.3.2 Two Evaluations

In the preceding analysis, the principal was not able to remember the agent's first period performance with certainty when he evaluated him at the end of the second period. In this section he makes an evaluation at the end of each period, so there will be no loss of information and, therefore, the parameter ρ does no longer appear. As each evaluation comes at fixed cost c_E, the total evaluation cost increases to $2c_E$. We will now calculate the conditions under which the agent chooses high effort. Furthermore, we will again derive the principal's optimal behavior and the corresponding profit. In a setting with two evaluations, the agent's utility over the two periods is:

$$E[U^t(e_1, e_2)] = \alpha^t + \beta^t f(e_1 + e_2) - (e_1 + e_2) = \alpha^t + (\beta^t f - 1)(e_1 + e_2)\,, \quad (3.9)$$

where α^t is the fixed wage component and β^t the bonus parameter for a contract with two evaluations. The principal's profit function now also takes into account the doubled evaluation cost:

$$E[\Pi^t(e_1, e_2)] = (1 - \beta^t) f(e_1 + e_2) - \alpha^t - 2c_E\,. \quad (3.10)$$

To derive the incentive compatible contract parameters in the case of two evaluations, we can use the results from Section 3.3.1 and set $\rho = 1$ as this corresponds to perfect memory. Using (3.3), (3.4), and with $\rho = 1$ we get $\beta^t = \frac{1}{f}$. Plugging this into equation (3.5) and again accounting for $\rho = 1$, we find the lowest α^t for this incentive compatible contract:

$$\alpha^t = \left[2 - 2\frac{f}{f}\right] e_H = 0\,. \quad (3.11)$$

These contract parameters allow us to calculate the principal's profit when he proposes a contract that implements high effort:

$$E[\Pi^t(e_H)] = \; 2(f-1)e_H - 2c_E. \quad (3.12)$$

It now remains to be determined under which parameter values the principal will prefer an incentive compatible contract to refraining from a contractual relationship. The condition for a contract to be profitable is:

$$E[\Pi^t(e_H)] \geq 0 \Leftrightarrow f \geq 1 + \frac{c_E}{e_H} := f^t\,. \quad (3.13)$$

If (3.13) does not hold, the principal will choose his outside option or evaluate only once.

3.3.3 Evaluating a Wealth-Constrained Agent

In the preceding analysis, we have implicitly assumed that the agent disposes over infinite wealth. No matter how high the fee the principal fixes in a contract with one evaluation, the agent will be able to pay this sum. In this subsection, we extend the analysis by allowing for a wealth-constrained agent. We do this by introducing a wealth bound $\bar{\alpha}$, which is the maximal amount the agent is able to pay in the first period. As long as in absolute terms $\bar{\alpha}$ is larger than or equal to the optimal fee α^o from Section 3.3.1 the analysis remains unaffected. But a restriction in the range $0 \geq \bar{\alpha} > \alpha^o$ has an impact on the contracts the principal can propose with one evaluation.[3] The analysis with two evaluations remains unchanged as, in this case, the optimal alpha is zero and, therefore, compatible with any wealth restriction.

To observe what happens in the case of a wealth restriction, we simply have to repeat the analysis for one evaluation, replacing the optimal α^o by $\bar{\alpha}$. This implies that the principal can only charge the agent an amount $\bar{\alpha}$, which is lower than the optimal one but he cannot lower incentives as he already chooses the lowest incentive compatible β. So a wealth constraint lowers the principal's profits. The participation constraint is tightened, whereas the incentive constraint remains unchanged. The expected profit for high effort in both periods with one evaluation becomes:

$$E[\Pi^o(e_H|\bar{\alpha})] = \quad (2f - \tfrac{1+\rho}{\rho})e_H - \bar{\alpha} - c_E \,. \tag{3.14}$$

To facilitate the interpretation, we can define the constant $a = \bar{\alpha} - \alpha^o$ which determines to what extent the magnitude of the wealth constraint differs from the optimal α^o in a contract with one evaluation. With this definition the wealth bound can be reformulated as $\bar{\alpha} = a - \frac{1-\rho}{\rho} e_H$. Using this expression, the princi-

[3]One could also introduce a minimum wage as constraint for α. So the restriction $\bar{\alpha} \leq 0$ is not crucial for the results, but has been chosen for reasons of consistency with the interpretation of the bound as a wealth-constraint.

pal's profit simplifies to

$$E[\Pi^o(e_H, a|\bar{\alpha})] = 2(f-1)e_H - a - c_E\,. \tag{3.15}$$

So one can see that the profit with one evaluation and a wealth constraint below zero is identical to equation (3.7) minus a, i.e. the deviation of the bound from the optimal fixed wage.

Returning to our previous notation, it is profitable to motivate the agent to exert high effort if:

$$E[\Pi^o(e_H|\bar{\alpha})] \geq 0 \Leftrightarrow f \geq \tfrac{1}{2}(\tfrac{1+\rho}{\rho} + \tfrac{\bar{\alpha}+c_E}{e_H}) := f^o(\bar{\alpha})\,. \tag{3.16}$$

This condition corresponds to the condition in (3.8), but due to the introduction of the wealth constraint $\bar{\alpha}$ the memory parameter ρ now also influences profitability. The principal can no longer entirely charge the agent for the cost of bounded memory, so his expected profit from a one-evaluation contract diminishes the closer ρ gets to zero. Due to $\bar{\alpha} \leq 0$ the profitability condition for two evaluations in (3.13) remains valid.

3.4 Results

We complete the analysis by deriving the behavior of the agent and the principal throughout the entire game. This is done by a comparison of the principal's profits which are summarized in Table 3.1. We will first deal with the case

Case:	$\alpha \gtreqless 0$	$\alpha \geq \bar{\alpha}$
$E[\Pi^o(e_H)]$	$2(f-1)e_H - c_E$	$(2f - \frac{1+\rho}{\rho})e_H - \bar{\alpha} - c_E$
$E[\Pi^t(e_H)]$	$2(f-1)e_H - 2c_E$	$2(f-1)e_H - 2c_E$

Table 3.1: All Profits

	One Evaluation	Two Evaluations
$E[\Pi^{o/t}(e_H)] \geq 0$	$f \geq 1 + \frac{c_E}{2e_H} = f^o$	$f \geq 1 + \frac{c_E}{e_H} = f^t$
$E[\Pi^{o/t}(e_H)]$	$2(f-1)e_H - c_E$	$2(f-1)e_H - 2c_E$
$\alpha^{o/t}$	$-\frac{1-\rho}{\rho}e_H$	0
$\beta^{o/t}$	$\frac{1}{f\rho}$	$\frac{1}{f}$

Table 3.2: Results for One and Two Evaluations

of unrestricted wealth. Without a binding wealth constraint, the profit with a single evaluation is always higher than with two evaluations, as the principal hands on the cost of his forgetfulness to the agent by "paying" a negative fixed compensation. In other words, switching from one to two evaluations leaves the principal's revenue unchanged—it is the agent who bears the entire cost of the principal's forgetfulness—and it only raises his fixed cost. Subsequently, we review this result when the agent's wealth is assumed to lie somewhere between the optimal α^o and zero. When the agent has a binding wealth constraint, two evaluations can be chosen for certain parameters of the model. The principal is no longer able to hand on the entire cost of his forgetfulness and hence, the two-evaluation contract becomes more advantageous.

No Wealth Constraint

Table 3.2 resumes critical values as well as optimal contract parameters and the corresponding profits in the unrestricted model. The critical values of the productivity parameter f, determining the principal's choice between the two contracts and the outside option, are always independent of ρ. For one evaluation, the more the principal forgets (smaller ρ), the more he will charge the agent via $\alpha^o = -\frac{1-\rho}{\rho}e_H$, where $\frac{\partial \alpha^o}{\partial \rho} = \frac{\rho-(\rho-1)}{\rho^2}e_H = \frac{1}{\rho^2}e_H > 0$. He can do so because he is a monopsonistic employer and the agent's outside option equals zero. The agent agrees as the fee he pays is outweighed by the variable compensation he receives. To put it another way, due to the principal's forgetfulness, the agent can extract a rent via the variable compensation. But thanks to his market power, the principal can recover the rent by charging a fee corresponding to the amount of the rent.

We determine the principal's choice concerning the evaluation frequency by a comparison of his possible gains with one and with two evaluations. Equation (3.17) illustrates a straight-forward result. As the principal can shift the cost of his forgetfulness to the agent, his profit with a single evaluation is always higher than with two evaluations since the latter imply a higher fixed cost:

$$\begin{aligned} & E[\Pi^o(e_H)] > E[\Pi^t(e_H)] \Leftrightarrow \\ & 2(f-1)e_H - c_E > 2(f-1)e_H - 2c_E \quad \Leftrightarrow \quad c_E > 0\,. \end{aligned} \tag{3.17}$$

This equation (3.17) leads us to the following proposition:

Proposition 1 *Evaluating once is a dominant strategy for the principal if he faces a risk-neutral agent who has no wealth constraint.*

Proof. As the evaluation cost is assumed to be larger than zero, equation (3.17) holds. Consequently, the profit from a one-evaluation contract exceeds the one of a two-evaluation contract. ■

Wealth-Constrained Agent

Comparing profits with one and with two evaluations leads to the following inequality:

$$E[\Pi^o(e_H|\bar{\alpha})] \gtreqless E[\Pi^t(e_H)] \Leftrightarrow -\tfrac{1-\rho}{\rho} e_H \gtreqless \bar{\alpha} - c_E\,. \tag{3.18}$$

While evaluating twice is unattractive as long as α is unbounded, this changes when we assume that the agent is wealth-constrained. Due to the fact that he is now unable to pay the optimal $\alpha^o < \bar{\alpha}$ for entering the contractual relationship, the extent of the principal's memory imperfection becomes determinant for his choice between the contracts. According to the inequality in (3.18), we can divide

Case:	$\alpha \gtreqless 0$	$\alpha \geq \bar{\alpha}$
$E[\Pi^o(e_H)] \gtreqless E[\Pi^t(e_H)]$	$c_E > 0$	$\alpha^o = -\frac{1-\rho}{\rho} e_H \gtreqless \bar{\alpha} - c_E$
$E[\Pi^o(e_H)] \geq 0$	$f \geq 1 + \frac{c_E}{2e_H}$	$f \geq \frac{1+\rho}{2\rho} + \frac{\bar{\alpha}+c_E}{2e_H}$
$E[\Pi^t(e_H)] \geq 0$	$f \geq 1 + \frac{c_E}{e_H}$	$f \geq 1 + \frac{c_E}{e_H}$
Result:	one evaluation or outside option	if $-\frac{1-\rho}{\rho} e_H > \bar{\alpha} - c_E$: one evaluations or outside option if $-\frac{1-\rho}{\rho} e_H < \bar{\alpha} - c_E$: two evaluations or outside option

Table 3.3: Tradeoffs between the Principal's Options

our results into two cases. In a first parameter range, the probability that the principal will forget the observed performance is high, relative to the difference of the fixed cost of employing the agent, i.e. the evaluation cost minus the fixed compensation component. Under these circumstances, the principal will either offer a contract with two evaluations inducing high effort in every period or he will not employ the agent. In a second parameter range, where the effect of the principal's imperfect memory is too weak for more than one evaluation to be profitable, the principal either chooses a contract with high effort and one evaluation or he does not offer any contract. Table 3.3 lists the tradeoffs between all relevant options of the principal. Table 3.4 reviews the optimal choices, critical values and outcomes when the agent is wealth-constrained.

Proposition 2 *If the agent has limited wealth, the optimal evaluation frequency depends on the following conditions:*

(i) *If* $-\frac{1-\rho}{\rho} e_H > \bar{\alpha} - c_E$, *the agent will be evaluated once.*

(ii) *If* $-\frac{1-\rho}{\rho} e_H < \bar{\alpha} - c_E$, *the agent will be evaluated twice.*

Otherwise, the principal is indifferent between one and two evaluations.

The proof directly follows from the inequality in (3.18).

	One Evaluation $\frac{1-\rho}{\rho}e_H < c_E - \bar{\alpha}$	Two Evaluations $\frac{1-\rho}{\rho}e_H > c_E - \bar{\alpha}$
$E[\Pi^{o/t}(e_H\|\bar{\alpha})] \geq 0$	$f \geq \frac{1+\rho}{2\rho} + \frac{\bar{\alpha}+c_E}{2e_H}$	$f \geq 1 + \frac{c_E}{e_H}$
$E[\Pi^{o/t}(e_H\|\bar{\alpha})]$	$(2f - \frac{1+\rho}{\rho})e_H - \bar{\alpha} - c_E$	$2(f-1)e_H - 2c_E$
$\alpha^{o/t}$	$\bar{\alpha}$	0
$\beta^{o/t}$	$\frac{1}{f\rho}$	$\frac{1}{f}$
$E[\Pi^{o/t}(e_H\|\bar{\alpha})] < 0$	$f < \frac{1+\rho}{2\rho} + \frac{\bar{\alpha}+c_E}{2e_H}$	$f < 1 + \frac{c_E}{e_H}$

Table 3.4: Results with a Wealth-Constrained Agent

3.5 A Setup with Depreciation of Output

In the previous analysis, we have described how a forgetful principal will design a linear contract when he might forget part of the agent's performance until the time of the evaluation, but takes the produced output entirely into account in his profit function. To point out the impact of the assumption that the principal benefits entirely from the output—no matter if he forgets it or not—we will now analyze how the optimal contract changes when the first period output enters the profit function with probability θ only, instead of being taken into account entirely. When θ equals ρ, the principal is not forgetful, but only evaluates the fraction of the output that remains after the depreciation. The case $\theta < \rho$ could be interpreted as a combination of depreciation and forgetfulness, while with $\theta > \rho$ we have a benevolent principal, who ignores part of the depreciation, determined by the difference of the two parameters. The agent's incentive and participation constraints in (3.3), (3.4) and (3.5) remain unaffected by this modification, but the expected profit for one evaluation now is denoted by:

$$E[\tilde{\Pi}^o(e_1, e_2)] = \theta\pi_1 + \pi_2 - \alpha^o - \beta^o(\rho\pi_1 + \pi_2) - c_E\,. \tag{3.19}$$

Plugging in the incentive compatible values for α^o and β^o yields:

$$E[\tilde{\Pi}^o(e_H)] = [\theta - \tfrac{1}{f\rho}\rho + 1 - \tfrac{1}{f\rho}]fe_H - \tfrac{\rho-1}{\rho}e_H - c_E\,. \tag{3.20}$$

Compared to equation (3.7), profit with one evaluation in this setup is lowered by $(1-\theta)fe_H$. As profit depends on the depreciation parameter θ, this parameter now influences profitability and, therefore, the choice of evaluation frequency. The profitability condition is:

$$\begin{aligned} &E[\tilde{\Pi}^o(e_H)] \geq 0 \Leftrightarrow \\ &f(1+\theta)e_H - 2e_H - c_E \geq 0 \Leftrightarrow \quad f \geq \tfrac{2e_H+c_E}{(1+\theta)e_H} := f^o(\theta)\,. \end{aligned} \tag{3.21}$$

To ensure profitability the productivity parameter f must be larger than the critical f^o we derived in (3.8). Furthermore, comparing profits with one and with two evaluations leads us to:

$$E[\tilde{\Pi}^o(e_H)] \gtreqless E[\Pi^t(e_H)] \Leftrightarrow f \lesseqgtr \frac{c_E}{(1-\theta)e_H}\,. \tag{3.22}$$

This result is different from the comparison in (3.17) as now there is a parameter range where the principal will prefer to evaluate an agent who is not wealth-constrained twice. To derive the parameter range where one evaluation is chosen, we combine the profitability condition and the condition for $E[\tilde{\Pi}^o(e_H, e_H)] \geq E[\Pi^t(e_H, e_H)]$. Then $f \geq \frac{2e_H+c_E}{(1+\rho)e_H}$ and $f < \frac{c_E}{(1-\rho)e_H}$ must hold at the same time. Such a productivity f exists if:

$$\frac{c_E}{(1-\theta)e_H} > \frac{2e_H + c_E}{(1+\theta)e_H} \Leftrightarrow \frac{1-\theta}{\theta}e_H < c_E\,. \tag{3.23}$$

The condition in (3.23) allows us to split our results into two categories, one where only two evaluations or the outside option will be chosen and another category where one or two evaluations will be chosen for profitable parameter ranges. The results are summarized in Table 3.5. As θ lowers the expected surplus for one evaluation, of course social welfare—in this setup this equals profit—is inferior to the one-evaluation welfare in a model without depreciation. Note that contrary to the results in Table 3.4, the critical condition does not split the results into a

	$\frac{1-\theta}{\theta}e_H > c_E$	$\frac{1-\theta}{\theta}e_H < c_E$
$\tilde{\Pi}^o(e_H) \geq 0$	-	$\frac{c_E}{(1-\theta)e_H} > f \geq \frac{2e_H+c_E}{(1+\theta)e_H}$
$\tilde{\Pi}^o(e_H)$	-	$f(1+\theta)e_H - 2e_H - c_E$
$\Pi^t(e_H) \geq 0$	$f \geq 1 + \frac{c_E}{e_H}$	$f > \frac{c_E}{(1-\theta)e_H}$
$\Pi^t(e_H)$	$2(f-1)e_H - 2c_E$	$2(f-1)e_H - 2c_E$
$\tilde{\Pi}^o(e_H)/\Pi^t(e_H) < 0$	$f < 1 + \frac{c_E}{e_H}$	$f < \frac{2e_H+c_E}{(1+\theta)e_H}$

Table 3.5: Results with Depreciation of Output

range with one evaluation and another with two evaluations, but for $\frac{1-\theta}{\theta}e_H < c_E$ both are possible, depending on the productivity f.

After doing the calculus, we would like to give some intuition for the modification of the profit function. Therefore we will shortly describe two examples that reflect the situation designed in the model. Both are related to research, one is about Ph.D. students, and another one about employees in R&D departments.

The classical way to obtain a Ph.D. used to be to write a thesis, consisting in a monograph, resuming the entire research work of several years, and a disputation of the monograph's content. Today, most Ph.D. students' work consists of writing papers which shall be published as soon as possible. While the classical Ph.D. student's research risks to be obsolete by the date of his disputation (sometimes years after the creation of the work), the up-to-date Ph.D. student, who is evaluated with each paper he finishes, has a clear advantage as his output is not subject to devaluation until the date of a possible publication. Of course, one could argue that the supervisor of a Ph.D. thesis that comes as a monograph should take into account that at the time of the research being actually done, this was state of the art. But one would at least expect that a supervisor will grade this student worse than a peer who's research output has not been devaluated, e.g. by similar but more recent publications on the same topic. Superiority of the newer practice could be suggested by its rapid propagation.

We can draw a similar scenario for an R&D department. Suppose researchers in the department are paid a bonus for patents providing the firm with a competitive advantage. If the evaluation of an initially successful patent occurs at a point in time where the competitive advantage has been destroyed by another more recent patent, this is likely to induce the supervisor to pay no bonus, as the patent does not provide an advantage for the firm at the time of the evaluation.

Both examples fit a model where output can lose its value over time and only the observed output at the time of the assessment enters the evaluation. With this in mind, it becomes intuitively clearer why in the basic model presented in Section 3.3.1 and 3.3.2 two evaluations must be a dominated choice. The principal has the entire market power, so he will naturally refrain from any action that diminishes welfare, i.e. his profit.

3.6 Conclusion

In this chapter, we analyzed which role frequency plays for performance evaluations if supervisors cannot retain information perfectly over time and agents are risk-neutral or wealth-constrained. The main result of the unrestricted model is as following. The fact that the quality of performance information depends on evaluation frequency does not automatically create a trade-off between evaluation and incentive cost. In a contract that contains variable and fixed compensation, the principal does not need more frequent evaluations of the agent to counterbalance his forgetfulness. The additional incentive cost caused by the loss of information due to imperfect memory will simply be shifted to the agent via a negative fixed compensation. Therefore, it is not surprising that the choice between providing high incentives and refraining from a contractual relationship remains unaffected as the level of the principal's forgetfulness varies.

But when we consider wealth-constrained agents, this outcome is weakened and now the principal's choice hinges on the degree of forgetfulness. More frequent evaluations become more profitable the higher the degree of forgetfulness, the lower the fixed cost of the evaluation, and the more the agent is wealth-constrained.

Due to the principal's monopsonistic position, his profit in the unconstrained model is identical to the entire surplus and the chosen contract leaves the agent with zero utility. With a binding wealth-constraint, the agent's utility is strictly positive for one evaluation. While the distribution of the surplus changes in this case, its size remains unaffected by the new split. But when first period incentives for one evaluation become too expensive the principal switches to two evaluations and lowers the surplus at stake due to an increased evaluation cost. Again, the agent's utility equals zero.

In our digression in Section 3.5, the evaluation frequency also affects the value of the first period output in the principal's profit function. This gives us more intuition why one evaluation is a dominant strategy in the basic model. We find examples from the domain of research for which the modified model seems particularly well-suited.

The way we set up the model also leaves space for further interpretations where supervisors face other types of bias. One example is systematical cautiousness. But the principal could also use what is referred to as a distorted performance measure (e.g. Moen and Rosén; 2005) and this distortion can be lowered at some cost. In addition, allowing for decreasing instead of fixed evaluation cost will strengthen the result under the wealth constraint. We expect more frequent evaluations to gain importance with respect to the basic model when we consider risk-averse agents. This is in line with the intuition given in Laffont and Martimort (2002), who assign similar features to situations with a wealth constraint as to a setup with risk aversion.

It seems plausible to assume that middle-managers and employees in comparable positions are wealth-constrained, though we cannot say anything about the extent of such a restriction. At least, we would expect them to be risk-averse. This suggests that for the design of optimal subjective review processes of middle-managers and alike, not only the measures used are important. Also, the frequency with which evaluations take place does play a role. While there is a strong focus on the nature of performance measures, up to now, there has been little research on the parameters determining the framework of the evaluation process. This chapter is a first step to investigate the interdependencies at work in this context. In the following chapter, we will deepen the analysis by explicitly introducing risk aversion of the agent in the model setup.

Chapter 4

Risk-Averse Agents

4.1 Introduction

In this chapter, we investigate which role the frequency of performance evaluations plays for incentive contracts when risk aversion is introduced. We present a straightforward model analyzing the impact of evaluation frequency in an employment relationship between a forgetful principal and a risk-averse agent. Within our two-period setting, we will derive how often a risk-neutral and forgetful employer should evaluate a risk-averse employee. Again, the employer reconsiders the agent's performance over the preceding period at periodical dates by attaching a subjective performance measure to the afore-gathered information. Due to bounded memory, the quality of the stored information varies depending on the length of the period the evaluation refers to. The principal has a rather precise idea on how well the agent performed in recent time, but events that are more remote are more likely to be forgotten. Each evaluation induces a fixed cost, such that a higher evaluation frequency is more costly on the one hand, but it leads to more precise information for the computation of a bonus payment on the other hand. Therefore, with each evaluation, the risk imposed on the agent is being reduced. Given the compensation parameters and the evaluation frequency chosen in the employment contract the agent decides in each period whether he will provide effort.

We find that risk aversion on behalf of the agent is at the origin of the principal's decision problem concerning the optimal evaluation frequency. If the agent was risk-neutral, evaluating only in the very last contractual period would be a dominant strategy. However, with a risk-averse agent, the optimal evaluation frequency will be determined by trading off the efficiency loss induced by the principal's bounded memory against the cost of an additional evaluation.[1] Furthermore, we can show in an extension that this trade-off—as well as its conditionality on the agent's attitude towards risk—is robust to the introduction of a second source of risk, i.e. a stochastic shock on output.

In Section 4.2, we present the setup of the model and its solution in a principal-agent structure. Section 4.3 discusses assumptions and possible extensions to the basic model in a general way. Section 4.4 sums up the results and concludes with the main insights on evaluation frequency we can extract from the model with a risk-averse agent.

4.2 The Basic Model with Risk Aversion

4.2.1 Setup

The basic model investigates the role of evaluation frequency in performance-based contracts by focussing on a particular time span, e.g. one year, in an ongoing employment relationship. To allow for inferences concerning evaluation frequency, this time span is subdivided into two periods. A risk-neutral principal is the owner of a production technology. To be able to extract a surplus from it, he must employ an agent who exerts a sufficiently high level of effort. For the

[1] This is in line with previous results in Chapter 3 where a wealth-constraint can be interpreted as a deviation from risk-neutrality (see Laffont and Martimort (2002, p. 155)).

two periods he designs an employment contract. This contract contains a linear compensation formula as well as a commitment to a frequency with which the agent will be evaluated during the contractual period (once or twice in our two-period model). Concentrating on linear compensation should not be seen as a restriction, as we analyze a discrete two-period model and, therefore, any type of payment structure can be represented by a linear formula. Another restriction deserves more attention here. We do not allow the principal to fully differentiate between the first and the second period in the contract, i.e. equal performance must be equally compensated for in both periods. In terms of a linear contract, this implies that for the two periods only one bonus factor determining the agent's share of the generated revenue can be fixed in the contractual agreement.[2] Once the principal has made a contract offer, the risk-averse agent will decide whether to work for him or not. In addition, he must decide on the levels of effort he is ready to exert in each of the periods.

The agent can choose between two possible levels of effort, e_L and e_H, which enter the production process and result in the principal's revenue $\pi_i = f \cdot e_i$ (with $i = H, L$). As for now, we do not consider a random shock, that could additionally influence the amount of the revenue.[3] The parameter $f > 1$ represents the productivity of the agent. We define $e_H, \pi_H > 0$ and normalize low effort and low output to zero, i.e. $e_L = \pi_L = 0$. The production function is common knowledge between the two contracting parties.

It is important to point out again that we analyze a situation with subjective performance evaluation. The employee's effort and the revenue—which is the value of the agent's output—are unverifiable for third parties but can be costlessly observed by the principal in each period. The agent dislikes effort, therefore the employer offers an incentive payment to compensate for the corresponding disu-

[2] As it will become clear that the results crucially depend on this assumption, we will extensively discuss it in Section 4.3.1.

[3] Results for a model version with a stochastic shock influencing the creation of the revenue are given in the appendix and will be discussed in Section 4.3.2.

tility e_i. The amount of this bonus is calculated as the product of a factor β, specified in the contract, and the information on the past performance the principal has at the last evaluation date. In addition to the bonus, the principal might want to include a fixed payment α into the contract.

Another important feature of the model is that the principal has bounded memory in the sense that he can forget observed performance information until the next period. In the model, ρ is the probability of the principal to remember the information and logically the probability of forgetting is $(1-\rho)$. Forgetting is possible because we consider a situation with subjective performance information which cannot be reconstructed later on if it has not been properly stored. It is only during the evaluation process that the employee's achievements are reviewed, quantified, made explicit and can therefore be preserved for later use. On the one hand, evaluating the agent prevents the principal from forgetting the observed level of performance but, on the other hand, this evaluation procedure comes at a cost c_E, which can be attributed to the time the principal spends on the analysis and treatment of the information necessary for the evaluation. As mentioned before, evaluations are truthful and will be carried out with certainty if this is specified in the contract.

Now, the agent's utility increases in fixed and variable pay and it decreases in effort. Due to the principal's bounded memory, a bonus will only be paid with probability ρ if previously there has been no evaluation. This is how the principal's forgetfulness introduces risk to the wage payment. As the agent is risk-averse, we assume a utility function that is overall concave. To keep the framework simple, effort cost and the fixed wage enter the agent's utility function linearly with their monetary value. Concavity comes through the first-period bonus payment that is valued only after the application of a strictly concave function $u(\cdot)$ to a bonus payment x. The latter function additionally has the properties $u(0) = 0$, and $u(x) < x \ \forall \ x$, where x is the total bonus payment of a period. It is not decisive for the results that bonus payments from different periods do not enter the utility function in the same way. This is the case as we are not interested in an inter-temporal comparison but in comparing two settings with different evaluation frequencies.

The chronology of events is analogous to the one in Chapter 3. When a positive surplus can be generated through the employment relationship, the principal designs a two-period work contract in period $\tau = 0$. The contract contains the number of times the agent will be evaluated, a fixed wage component α, and a bonus factor β. This bonus factor determines variable compensation when it is multiplied with the level of performance of the agent, which has been assessed in the evaluation. If one party rejects the contract, both players receive the utility of their outside option. After the acceptance of the contract, the employee exerts effort $e_1 \in \{e_L, e_H\}$ in period $\tau = 1$ to generate the first-period revenue $\pi_1 \in \{\pi_L, \pi_H\}$. Also, the fixed wage α is payed in advance.[4] Analogously, in period $\tau = 2$ the agent exerts effort $e_2 \in \{e_L, e_H\}$ to create $\pi_2 \in \{\pi_L, \pi_H\}$. Now, the expected bonus payment is calculated and payed to the agent. The computation of the bonus depends on the evaluation frequency. When the principal evaluates only in the second period, he remembers the first-period revenue $\pi_1 = f \cdot e_1$ with probability $\rho \in]0, 1[$. A fixed evaluation cost of c_E arises, and the agent receives an expected bonus of $\beta \cdot (\rho\pi_1 + \pi_2) = \beta f(\rho e_1 + e_2)$ at the end of the second period. With two evaluations, the principal reviews the employee's performance at the end of each period and, therefore, correctly remembers what he has observed over the entire time span. His total evaluation cost increases to $2c_E$ and the agent receives a bonus $\beta(\pi_1 + \pi_2) = \beta f(e_1 + e_2)$ which accurately reflects his performance during the two periods. Note that the bonus corresponding to the low effort level equals zero.

We solve two versions of the discrete model, a version with one evaluation, where the agent risks not to be payed the first-period bonus, and a second ver-

[4]One should keep in mind that in a framework without discounting, the timing of the payments is not very important. For example, paying out the fixed compensation during the first period is equivalent to paying $\alpha' = \frac{\alpha}{2}$ in every period.

sion with two evaluations. For each model version, backward induction yields the subgame-perfect equilibrium. First, we calculate for which compensation parameters α and β the agent is ready to join the contract and to work hard. With this information at hand, the principal maximizes his profit by choosing the lowest payment inducing high effort in both periods. Comparing the outcomes with one and with two evaluations allows us to make inferences on the optimal evaluation frequency. Although we present a two-period model, we do not consider discounting in order to focus entirely on the impact of imperfect memory on optimal evaluation frequency.

4.2.2 One Evaluation

When designing the contract, the principal has two strategies at hand. He can choose to evaluate once despite the fact that this imposes a risk on the agent or he prefers to evaluate twice. To model his choice, we first specify the parties' payoffs and then compare them under the different regimes. We start with the case of one evaluation where the principal might understate the agent's first period performance when he evaluates him at the end of the second period. Expressed in monetary terms, the agent's utility for high effort over the entire time span can be written as the sum of expected utilities in each period:

$$E[U^o(e_1, e_2)] = E[U_1^o(e_1)] + E[U_2^o(e_2)] = \alpha^o + \rho u(\beta^o f e_1) - e_1 + \beta^o f e_2 - e_2\,, \quad (4.1)$$

where $u(\cdot)$ is the concave function introducing the employee's risk aversion. An employment contract is incentive-compatible if, in every period, working hard is at least as good for the agent as not doing so.[5] Therefore, the incentive constraints

[5] We do not account for situations where the principal would require different effort levels for the two periods. This would considerably complicate an analysis in a discrete model due to the occurrence of more cases, but it would not create substantially more insights.

for high effort in Period 1 and 2 are:

$$E[U_1^o(e_H)] \geq E[U_1^o(e_L = 0)] \Leftrightarrow \beta^o \geq \frac{1}{fe_H} u^{-1}\left(\frac{e_H}{\rho}\right) , \tag{4.2}$$

and

$$E[U_2(e_H)] \geq E[U_2(e_L = 0)] \Leftrightarrow \beta^o f e_H - e_H \geq 0 \Leftrightarrow \beta^o \geq \frac{1}{f} . \tag{4.3}$$

The participation constraint postulates that, in expectancy, the utility over the entire time span must be at least as large as the utility the agent can derive from his outside option which is $\bar{U} = 0$:

$$E[U^o(e_H)] \geq E[\bar{U}] \Leftrightarrow \alpha^o + \beta^o f e_H + \rho u(\beta^o f e_H) - 2e_H \geq 0 . \tag{4.4}$$

We restricted the contractual agreement such that the principal has one and the same bonus factor at hand to steer incentives over the two periods. Hence, we must identify which of the constraints in (4.2) and (4.3) is binding. As in a contract with one evaluation, it is the agent's first-period performance the principal is likely to forget, he must set higher incentives for effort exertion in the first period. So (4.2) is the binding incentive constraint, while (4.3) is slack. Using (4.2), we can now choose the incentive-compatible bonus factor for which the variable wage bill is the lowest, i.e. $\beta^{o*} = \frac{1}{fe_H} u^{-1}(\frac{e_H}{\rho})$. One can easily show that this bonus factor automatically complies with (4.3), as $u^{-1}(\frac{e_H}{\rho}) > u^{-1}(e_H)$ holds. Given the optimal β^{o*}, the profit maximizing principal will pay the lowest fixed wage α^o that fulfills the participation constraint in (4.4):

$$\alpha^{o*} = e_H - u^{-1}(\tfrac{e_H}{\rho}) . \tag{4.5}$$

This optimal fixed wage component is negative, which might seem counterintuitive at first sight. Why should a risk-averse agent be willing to enter a contract that is not only risky but also requires some kind of entry fee? But the negativeness of α^{o*} can be easily explained by the asymmetry between the two contracting periods. The principal steers incentives via one bonus factor, though the periods are not identical. Due to the principal's bounded memory and the concavity in the first-period bonus payment, the agent needs higher incentives to exert effort in the first than in the second period. However, this implies that incentives for the second period are too high, the agent is paid more than necessary to be willing to choose e_H. It is this rent that the monopsonistic principal reappropriates when choosing a negative fixed compensation parameter. In a setup where the employee's outside option is larger than zero this would imply that an employer who must pay inefficiently high bonuses to motivate his employees will decide to cut the fixed wage in turn. With one evaluation, the principal's expected profit is the output over the two periods minus the agent's compensation and minus the evaluation cost:

$$E[\Pi^{o*}(e_H)] = (2f-1)e_H - \rho u^{-1}\left(\frac{e_H}{\rho}\right) - c_E \,. \tag{4.6}$$

Under a one-evaluation contract, the agent's expected compensation is composed of the fixed payment plus the bonus for the level of performance that is determined in the evaluation at the end of the second period. Using the incentive-compatible α^{o*} and β^{o*} we have derived above, we can specify the principal's expected wage bill as:

$$E[C^{o*}(e_H)] = e_H + \rho u^{-1}\left(\frac{e_H}{\rho}\right) \,. \tag{4.7}$$

Comparing the expected compensation to the agent's disutility of effort, which is $2e_H$, leads us to the following result:

Proposition 3 *The monetary value of the wage payment strictly exceeds the agent's disutiltiy of effort, i.e.* $E[C^{o*}(e_H)] = e_H + \rho u^{-1}(\frac{e_H}{\rho}) > 2e_H$.

Proof. The condition comparing (6.9) to $2e_H$ can be simplified as following:

$$e_H + \rho u^{-1}\left(\frac{e_H}{\rho}\right) \gtreqless 2e_H \Leftrightarrow u^{-1}\left(\frac{e_H}{\rho}\right) \gtreqless \frac{e_H}{\rho} \Leftrightarrow \frac{e_H}{\rho} \gtreqless u\left(\frac{e_H}{\rho}\right). \tag{4.8}$$

As we have assumed a strictly lower utility for risky payments than for non-risky ones, i.e. $u(x) < x \ \forall \ x$, it directly follows that $\frac{e_H}{\rho}$ is larger than $u(\frac{e_H}{\rho})$ and, therefore, Proposition 3 holds. ■

The principal has the entire market power, and hence he will pay the agent a risk premium which sets the expected utility of the agent equal to zero:

$$E[U^{o*}(e_H)] = e_H - u^{-1}\left(\frac{e_H}{\rho}\right) + u^{-1}\left(\frac{e_H}{\rho}\right) + e_H - 2e_H = 0\,. \tag{4.9}$$

In expectations, offering a one-evaluation contract is profitable to the employer as long as:

$$E[\Pi^o(e_H)] \geq 0 \Leftrightarrow f \geq \frac{1}{2e_H}\left[\rho u^{-1}\left(\frac{e_H}{\rho}\right) + c_E + e_H\right] := \ f^o\,. \tag{4.10}$$

If the agent's productivity f does not fulfill the inequality in (4.10), the principal will either choose his outside option or evaluate twice. The threshold f^o will be used later on to characterize competitiveness within the extension to a duopolistic model in Section 5.2.

4.2.3 Two Evaluations

Having calculated the relevant payoffs for one evaluation, we will now analyze the outcome of a two-evaluation contract between the principal and his risk-averse employee. As the principal evaluates at the end of each period, there will be no loss of information in the evaluation process. Hence, the parameter ρ no longer plays a role in the equilibrium contract. But since each evaluation comes at a fixed cost c_E, the total evaluation cost increases to $2c_E$. Analogous to Section 4.2.2, we first calculate the conditions under which the agent chooses high effort in a two-evaluation contract. Subsequently, we will derive the principal's optimal behavior and the corresponding profit. In a setting with two evaluations, the agent's expected utility over the two periods becomes:

$$E[U^t(e_1, e_2)] = \alpha^t + u(\beta^t f e_1) - e_1 + \beta^t f e_2 - e_2 \,. \tag{4.11}$$

The principal's profit function with two evaluations is composed of his share in the generated revenue minus the fixed compensation and the doubled evaluation cost:

$$E[\Pi^t(e_1, e_2)] = (1 - \beta^t) f(e_1 + e_2) - \alpha^t - 2c_E \,. \tag{4.12}$$

To derive the incentive-compatible contract parameters in the case of two evaluations, we can use the incentive and participation constraints from Section 4.2.2. It suffices to change the memory parameter ρ. Setting it equal to one reflects that the principal has perfect memory, as we have assumed it for the setup with two evaluations. From condition (4.2), we can analogously derive the optimal bonus factor $\beta^{t*} = \frac{1}{fe_H} u^{-1}(e_H)$. We insert this lowest incentive-compatible bonus parameter into (4.4) to find the smallest α^t for which the participation constraint is binding:

$$\alpha^{t*} = e_H - u^{-1}(e_H) \,. \tag{4.13}$$

Given the optimal contract parameters, the principal's profit for a contract that implements high effort becomes:

$$E[\Pi^{t*}(e_H)] = (2f - 1)e_H - u^{-1}(e_H) - 2c_E \,. \tag{4.14}$$

Finally, the following condition reflects for which parameter values it is profitable for the principal to offer an incentive-compatible contract with two evaluations:

$$E[\Pi^t(e_H)] \geq 0 \Leftrightarrow f \geq \frac{1}{2e_H}(u^{-1}(e_H) + e_H + 2c_E) := f^t \,. \tag{4.15}$$

Analogous to condition (4.10), the principal will either choose his outside option or evaluate only once if inequality (4.15) does not hold. Together with the value f^o defined by (4.10), f^t will be used as a threshold to define under which circumstances the heterogenous principals in Section 5.2 face a competitive situation.

4.2.4 Results and the Impact of Risk Aversion

Knowing the principal's behavior for each evaluation frequency, it is now possible to determine when it is preferable for the principal to evaluate once—despite the fact that this imposes a risk on the agent—and when he wants to evaluate twice. The following critical condition arises from a comparison of the profit with one evaluation to the profit with two evaluations:

$$E[\Pi^{o*}(e_H)] \gtreqless E[\Pi^{t*}(e_H)] \Leftrightarrow c_E \gtreqless \rho u^{-1}\left(\frac{e_H}{\rho}\right) - u^{-1}(e_H) = \bar{c}_E\,. \tag{4.16}$$

From this condition, we cannot directly derive whether the right hand side is strictly larger than zero, i.e. whether there exists a trade-off between the advantages of an additional evaluation and the strictly positive cost this evaluation entails. But if we use, for example, a logarithmic specification, i.e. $u(x) = ln(x)$, we find that the right hand side of the inequality exceeds zero as long as ρ is smaller than one and when values of e_H are sufficiently high. The essential reason for such a trade-off to exist is the agent's attitude towards risk. Let us assume the agent was risk-neutral and his utility from the bonus payment was entirely linear, i.e. $u(x) = x$. It is easy to show that in this case the threshold $\bar{c}_E$ from the condition in (4.16) equals zero. As the evaluation cost is assumed to be strictly larger than zero, evaluating twice is never optimal when the agent is risk-neutral.[6] Table 4.1 gives an overview of the outcomes chosen by the princi-

[6]This is in line with the results in Chapter 3 where two evaluations are always a dominated strategy.

	One Evaluation $c_E > \bar{c}_E$	Two Evaluations $c_E < \bar{c}_E$
$\alpha^{o/t}$	$e_H - u^{-1}(\frac{e_H}{\rho})$	$e_H - u^{-1}(e_H)$
$\beta^{o/t}$	$\frac{1}{fe_H}u^{-1}(\frac{e_H}{\rho})$	$\frac{1}{fe_H}u^{-1}(e_H)$
$E[\Pi^{o/t}(e_H)]$	$(2f-1)e_H - \rho u^{-1}(\frac{e_H}{\rho}) - c_E$	$(2f-1)e_H - u^{-1}(e_H) - 2c_E$
$E[\Pi^{o/t}(e_H)] \geq 0$	$f \geq \frac{1}{2e_H}(\rho u^{-1}(\frac{e_H}{\rho}) + e_H + c_E)$	$f \geq \frac{1}{2e_H}(u^{-1}(e_H) + e_H + 2c_E)$

Table 4.1: Results with a Risk-Averse Agent

pal and the corresponding values of the compensation variables. Depending on the critical condition in (4.16), the employer will choose between one and two evaluations. The first two lines of Table 4.1 give the optimal values of the compensation variables for different evaluation frequencies. The third line contains the corresponding profits which the principal can realize. Finally, only if at least one of the profitability constraints, (4.10) and (4.15), in the last line is fulfilled, the principal will decide to employ the agent.

To get a more precise idea on the implications that the agent's risk aversion has on contract design, we would also like to know how the compensation parameters differ from the results with a risk-neutral agent in Chapter 3. In the following inequalities, the subscript RN stands for values from a setup where the agent is risk-neutral and RA stands for the case of a risk-averse agent. Due to $u(x) < x$, the principal's expected profit for one evaluation is strictly lowered by the introduction of risk aversion on part of the agent:

$$E[\Pi^o_{RA}(e_H)] < E[\Pi^o_{RN}(e_H)] \Leftrightarrow u\left(\frac{e_H}{\rho}\right) < \frac{e_H}{\rho}\,. \tag{4.17}$$

For the same reason, variable compensation is larger under risk aversion:

$$\beta^o_{RA} > \beta^o_{RN} \Leftrightarrow u\left(\frac{e_H}{\rho}\right) < \frac{e_H}{\rho}\,, \tag{4.18}$$

whereas fixed compensation diminishes:

$$\alpha^o_{RA} < \alpha^o_{RN} \Leftrightarrow u\left(\frac{e_H}{\rho}\right) < \frac{e_H}{\rho}\,. \tag{4.19}$$

Analogous effects arise for a two-evaluation contract where the corresponding inequalities are

$$E[\Pi^t_{RA}(e_H)] < E[\Pi^t_{RN}(e_H)] \Leftrightarrow u(e_H) \gtreqless e_H\,, \tag{4.20}$$

$$\beta^t_{RA} > \beta^t_{RN} \Leftrightarrow u(e_H) < e_H\,, \tag{4.21}$$

and

$$\alpha^t_{RA} < \alpha^t_{RN} \Leftrightarrow u(e_H) < e_H\,. \tag{4.22}$$

Thus, we can conclude that the introduction of the agent's risk aversion raises variable compensation whereas fixed compensation is lowered and this is independent of the evaluation frequency. At first sight, this result does not seem intuitive. One would expect lower incentives and higher fixed compensation for a risk-averse agent. But as the discrete model allows for two effort levels only, i.e. $e_L = 0$ and $e_H > 0$, the agent cannot adjust optimally by reducing effort and it is the principal who has to adjust incentives to the given effort levels. Due to the intertemporal asymmetry, he will also adjust the fixed wage α downwards.

The inequalities in (4.17) and (4.20) show that, compared to a contract with a risk-neutral agent, the principal's profit diminishes by a risk premium when he faces a risk-averse agent. This complies with the standard result that risk aversion on the part of the agent lowers efficiency.[7] We resume the results concerning the impact of risk aversion in the following proposition:

[7] The criterion for efficiency in this setting is to maximize the expected surplus $U^o + \Pi^o$, which equals the expected profit when the principal is modeled as a monopsonist on the labor market.

Proposition 4 *In a contract with one (two) evaluation(s), the introduction of risk aversion on behalf of the agent leads to the following effects:*
a) $\beta^{o/t} \uparrow$ $(\beta^{o/t}_{RA} > \beta^{o/t}_{RN})$*, i.e. the principal sets higher incentives;*
b) $\alpha^{o/t} \downarrow$ $(\alpha^{o/t}_{RA} < \alpha^{o/t}_{RN})$*, i.e. the agent receives a lower fixed compensation;*
c) $E[\Pi^{o/t}] \downarrow$ $(E[\Pi^{o/t}_{RA}(e_H)] < E[\Pi^{o/t}_{RN}(e_H)])$*, i.e. the principal's profit which equals the surplus is diminished.*

Beyond the effects induced by risk aversion, we are interested in the impact of variations in the memory parameter ρ on the compensation parameters in the case of one evaluation. Comparative statics confirm that α_{RA} rises with better memory and the bonus factor β_{RA} diminishes:

$$\frac{\partial \alpha_{RA}}{\partial \rho} = \frac{e_H}{\rho^2} u_\rho^{-1}\left(\frac{e_H}{\rho}\right) > 0 \tag{4.23}$$

$$\frac{\partial \beta_{RA}}{\partial \rho} = \frac{1}{f e_H} \frac{\partial (u^{-1}(\frac{e_H}{\rho}))}{\partial \rho} = -\frac{1}{f\rho^2} u_\rho^{-1}\left(\frac{e_H}{\rho}\right) < 0 \tag{4.24}$$

Both conditions hold because $u(x)$ is an increasing function and, therefore, $u^{-1}(\cdot)$ is, too.

$$\tfrac{\partial \Pi_{RA}}{\partial \rho} = -u^{-1}(\tfrac{e_H}{\rho}) - \rho\left(-\tfrac{e_H}{\rho^2} u_\rho^{-1}(\tfrac{e_H}{\rho})\right) \tag{4.25}$$

Finally, it is quite intuitive and one can easily check that the efficiency gains due to a marginal change in ρ correspond to the marginal decrease in the agents wage payment: $\frac{\partial \Pi^o_{RA}}{\partial \rho} = -\frac{\partial C^o_{RA}}{\partial \rho}$.

Proposition 5 *Independent of the agent's level of risk aversion, a rise in the memory parameter* ρ*, i.e. a better memory capacity of the principal, has the following effects in a contract with one evaluation:*
a) $\beta^o \downarrow$*, i.e. the principal sets lower incentives when* ρ *rises;*
b) $\alpha^o \uparrow$*, i.e. the agent receives a higher fixed wage.*
When the agent is risk-neutral, the expected profit $E[\Pi^o]$ *remains unaffected by changes in* ρ*, whereas it is strictly lowered when the agent is risk-averse.*

4.3 Discussion and Extensions

In this section we will discuss important assumptions in the basic model and how some of them can be justified by possible extensions of the setup. The extensions we propose are the integration of a stochastic shock on output (Section 4.3.2) and an analysis of the employment contract when the principal as well as the agent face an infinite time horizon (Section 4.3.3).

4.3.1 Critical Assumptions

The first assumption that might be subject to criticism, is that forgetting means more precisely 'forgetting good performance' whereas we never allow the principal to remember, so to say to invent, something which has not happened. Of course, concentrating on the probability of the principal to remember the output is particularly convenient for modeling. We admit that this is a rather simplistic way to describe the complex functioning of the human memory but, at the same time, we believe that allowing for something like "hallucinations", i.e. "remembering" things that never happened, includes very complicated mechanisms of the mind that should be modeled with a lot of care. When could an employer remember good performance that has never occurred? This is only plausible if the employment history of the agent or the impression the employer has got of his employee suggest that the latter is a good performer. Taking such a pigeonhole effect into account would add a lot of complexity to the model and even more assumptions would have to be justified. Though, it does not capture a bias due to prejudices of the principal concerning the quality of the agent's performance, our simple framework gives us at least some first insights into how evaluation frequency affects incentives if the agent is uncertain whether his performance will be rewarded or not. Furthermore, interpreting the cause of this uncertainty as bounded memory is only one possibility. The simplicity of the introduction of the bias leaves space for alternative interpretations, which will be discussed in the conclusion.

A second critical assumption in the basic model refers to the bonus factor β. We restrict the analysis to contracts where β is equal over the two periods. As it has already been pointed out to some extent throughout the preceding sections, this drives a lot of our results. Certainly, this assumption is particularly strong as contractual freedom should allow the principal to differentiate between the periods. But on the other hand, it is also plausible to argue that contractual complexity will be considerably enlarged when allowing for different bonus parameters. However, we would like to bring forward another argument sustaining that the restriction is reasonable. What the model actually does, is allowing the principal to choose one bonus parameter per evaluation period, which comprises either one or two periods. Although he could pick different β^t in a contract with two evaluations he would never do so. The two periods are identical and, therefore, the optimal β^{t*} are as well. So altogether, we do not generally prohibit a differentiation between periods, but we say that a differentiation of the bonus factors with respect to the evaluation periods is rather unrealistic due to contractual complexity. And this is one reason why fewer evaluations do not only lower evaluation cost, but also raise the cost of providing incentives.

4.3.2 Model with a Stochastic Shock

A model version incorporating moral hazard can be found in the appendix to this chapter. In this modification, we assume that the principal can only observe the output generated by the agent which is affected by an unobservable stochastic shock. As the extension gives us relatively few additional insights, we will only discuss some main results at this point. Some of the results are quite straightforward, e.g the fact that the bonus of the agent must increase—the evaluation frequency being irrelevant—compared to the setup without the shock. Not only bounded memory is a risk for his bonus payment, but also the output might not even materialize and, consequently, with some positive probability no bonus will be paid at all, even though the employee has put forth the required level of effort.

With an additional source of risk, it becomes even harder to make a valid statement on the optimal evaluation frequency. For a logarithmic specification of the utility function, one can show that a one-evaluation contract does not dominate the two-evaluation contract regardless of the evaluation cost. Depending on the parametrization of the model and the shape of the agent's utility function, evaluations will be more or less frequent with the stochastic shock than without its occurrence.

Two effects are at work and their extent depends on the shape of the utility function as well as the chosen parameters. First, the possibility that there is no revenue, despite the fact that the agent has exerted effort, lowers the value of the evaluation. The principal cannot differentiate between situations in which the agent has not exerted effort and in which there is no output due to the shock. Therefore, evaluations become less important, because now the event where the principal forgets to pay a bonus occurs with a lower probability. The probability for this state to emerge is the joint probability of two events. High output must realize and at the same time, the principal must forget that this event has happened. Second, compared to the setup with a deterministic production technology, the optimal bonus payment is larger with the stochastic shock. Thus, if a revenue is generated and the principal forgets to pay the bonus, the damage for the agent is large.

4.3.3 Infinite Horizon

So far, we have assumed that the employer will stick to the contract he has offered, although some information in this contract is not verifiable for third parties, especially not for a court. In economic modeling, for such an implicit contract to be stable, the expected gain from the persistence of the contractual relationship must be large enough to prevent the employer from cheating on the agent. Cheating here means a deviation from the terms of the implicit contract, e.g. paying a lower or no bonus and skipping evaluations specified in the contract. Formally, a stable implicit contract is characterized by a situation where the expected gain from a one-time deviation from the implicit contract $E[\Pi_D]$ is equal or lower than the expected gain from sticking to what has been agreed upon, $E[\Pi_{IC}]$, now and in the future. This can be summarized in the following stability condition:

$$E[\Pi_D] \leq \left(1 + \frac{1}{r}\right) E[\Pi_{IC}] \Leftrightarrow r \leq E\left[\frac{\Pi_{IC}}{\Pi_D - \Pi_{IC}}\right] := r^{crit}\,, \tag{4.26}$$

where r is the employer's interest rate. If he discounts strongly due to a high interest rate, the termination of the beneficiary employment relationship with the agent does not prevent the principal from deviating and exploiting the agent. The same is true for high gains from this deviation, i.e. $\Pi_D - \Pi_{IC}$ is large.

Depending on the evaluation frequency that has been chosen, one can select the proper $E[\Pi_{IC}]$ from equation (4.6) or (4.14). It is harder to select an appropriate value for the expected gain from a deviation. Hence, we will choose two extreme scenarios here, where evaluations are not contractible and no bonuses will be paid. So when he deviates, the principal will skip the last evaluation, as it is costly. In a two-evaluation contract, he will still conduct an evaluation after the first period, otherwise the agent would become suspicious and stop working hard after the first period.[8] If one of the following two critical values is larger than the principal's interest rate r, there exists an implicit contract that is stable

[8]More possible scenarios for deviations on behalf of the principal are discussed in Section 4.5.2 in the Appendix to this chapter.

for the corresponding evaluation frequency:

$$\begin{aligned} r_2^o &= \frac{2fe_H - e_H - \rho u^{-1}(\frac{e_H}{\rho}) - c_E}{(1+\rho)u^{-1}(\frac{e_H}{\rho}) + c_E}, \\ r_2^t &= \frac{2fe_H - e_H - u^{-1}(e_H) - 2c_E}{2u^{-1}(e_H) + c_E}. \end{aligned} \tag{4.27}$$

Obviously, the two thresholds are strictly positive for beneficial relationships. A rise in the evaluation cost clearly lowers the value of the thresholds, so implicit contracting becomes more difficult.

4.4 Conclusion

In the present chapter, a principal with bounded memory faces a trade-off concerning the choice of evaluation frequency when the agent is risk-averse. Depending on the cost of an additional evaluation, it might be worthwhile for an employer to increase evaluations as this allows to lower variable compensation. Whereas, in a model with a risk-neutral agent (Chapter 3) more frequent evaluations are superfluous—they simply increase the fixed cost without yielding any benefit—in a model with risk aversion more evaluations can even enhance efficiency through the elimination of risk within the evaluation process. By absorbing the effect of the principal's memory imperfection, they eliminate the efficiency loss due to risk imposed on the agent.

The existence of the trade-off is relatively robust to a setup featuring a stochastic shock on output. Whether a principal should evaluate more or less often when there is a shock crucially depends on the specification of the model. Two opposite effects are at work. On the one hand, the possibility that there is no revenue, despite the agent exerting effort, lowers the value of the evaluation. On the other hand, the bonus payment is larger in the setup with the stochastic shock. Consequently, if a revenue is generated and the principal forgets to pay the bonus, the damage for the agent is larger than without the stochastic shock.

The performance measure we use in the model can be considered a subjective measure. In Section 4.3.3, we discuss conditions for the stability of the optimal contracts we propose, based on the analysis of the basic model. Stability of the implicit contracts certainly depends on the parameter choice. We find a clear negative impact of the evaluation cost on stability, whereas the effect of other variables depends on the specification of the model.

4.5 Appendix

4.5.1 Basic Model with a Stochastic Shock

Up to now, if the agent exerted high effort, this automatically resulted in high output, i.e. $\pi(e_H) = f \cdot e_H = \pi_H$. In this section, the model will be enriched by a stochastic shock. High effort only improves the agent's chance to produce high output. Due to a stochastic shock, a hard-working agent can be unlucky and produce low output. Likewise, a lazy agent can be lucky and produce high output without having exerted effort. This is modeled in the following way: If an agent exerts high effort, he will generate a high benefit π_H with probability $p(e_H) = p_H$ and a low benefit $\pi_L = 0$ with probability $(1 - p_H)$. Low effort, i.e. $e_L = 0$, generates the high benefit π_H only with probability $p(e_L) = p_L < p_H$ and, therefore, implies a larger probability $(1 - p_L)$ to produce $\pi_L = 0$. Whenever he generates $\pi_H > 0$, the agent is paid a bonus, no matter the effort he has shown. Again, we will compare a situation where the principal evaluates only once with a setup where there are two evaluations.

One Evaluation

The agent's expected utility from the contract is

$$E[U^o(e_1, e_2)] = \alpha^o + \rho p(e_1)u(\beta^o \pi_H) - e_1 + p(e_2)\beta^o \pi_H - e_2\,. \tag{4.28}$$

Comparing his payoffs with high and low effort respectively yields the following incentive compatible bonus factors:

$$E[U_1^o(e_H)] \geq E[U_1^o(e_L = 0)] \Leftrightarrow \beta^o \geq \frac{1}{\pi_H} u^{-1}\left(\frac{k}{\rho}\right) \tag{4.29}$$

and

$$E[U_2(e_H)] \geq E[U_2(e_L = 0)] \Leftrightarrow \beta^o \geq \frac{1}{\pi_H} k\,, \tag{4.30}$$

where $k := \frac{e_H}{p_H - p_L}$. To ensure that the agent prefers working to his outside option

$$E[U^o(e_H)] \geq E[\bar{U}] \Leftrightarrow \alpha^o + \rho p_H u(\beta^o \pi_H) + p_H \beta^o \pi_H - 2e_H \geq 0 \tag{4.31}$$

must equally hold. As we allow only for one bonus factor, the principal must choose the factor resulting from (4.29) to ensure hight effort in both periods. So, this first condition is tight, whereas (4.3) is slack. A contract offer that fulfills

$$\alpha^o \geq 2e_H - p_H \left[k + u^{-1}(\frac{k}{\rho})\right] \tag{4.32}$$

can ensure the participation of the principal. Given the optimal contract parameters α^{o*} and β^{o*}, the principal's expected profit equals

$$E[\Pi^{o*}(e_H)] = 2(p_H \pi_H - e_H) + p_H \left[k - \rho u^{-1}\left(\frac{k}{\rho}\right)\right] - c_E\,. \tag{4.33}$$

The agent's expected compensation amounts to

$$E[C^{o*}(e_H)] = 2e_H + p_H \left[\rho u^{-1}\left(\frac{k}{\rho}\right) - k\right] \tag{4.34}$$

and outweighs the monetary value of his effort. But due to his risk aversion, his expected utility equals zero:

$$E[U^{o*}(e_H)] = 2e_H - p_H\left[k + u^{-1}\left(\frac{k}{\rho}\right)\right] + p_H\left[k + u^{-1}\left(\frac{k}{\rho}\right)\right] - 2e_H = 0\,. \quad (4.35)$$

Finally, it is important to point out that there exists a critical threshold for the level of revenues that must be respected for the employment relationship to be profitable in a one-evaluation setup:

$$E[\Pi^o(e_H)] \geq 0 \Leftrightarrow \pi_H \geq \tfrac{1}{2p_H}\left(2e_H + p_H\left[\rho u^{-1}\left(\tfrac{k}{\rho}\right) - k\right] + c_E\right) := \ f^o. \quad (4.36)$$

Two Evaluations

With two evaluations, the agent's utility can be described by

$$E[U^t(e_1, e_2)] = \alpha^t + p(e_1)u(\beta^t\pi_1) - e_1 + p(e_2)\beta^t\pi_2 - e_2\,, \quad (4.37)$$

whereas the principal's expected profit is

$$E[\Pi^t(e_1, e_2)] = 2p_H\pi_H - \alpha^t - (p(e_1) + p(e_2))\beta^t\pi_H - 2c_E\,. \quad (4.38)$$

The optimal bonus factor β^{t*} can be calculated from (4.29) by setting $\rho = 1$ and the following α^{t*} ensures the agent's participation:

$$\alpha^{t*} = 2e_H - p_H[k + u^{-1}(k)]\,. \quad (4.39)$$

The equilibrium profit hence amounts to

$$E[\Pi^{t*}(e_H)] = 2p_H\pi_H - 2e_H + p_H[k - u^{-1}(k)] - 2c_E\,. \quad (4.40)$$

For the contract to be signed, this profit must be larger than zero in expectations:

$$E[\Pi^t(e_H)] \geq 0 \Leftrightarrow \pi_H \geq \frac{1}{2p_H}\left[2e_H + p_H[u^{-1}(k) - k] + 2c_E\right] := f^t\,. \tag{4.41}$$

Results and the Impact of Risk Aversion

The introduction of a second source of risk, namely a stochastic shock on output, has an impact on the trade-off between one and two evaluations:

$$\begin{gathered} E[\Pi^{o*}(e_H)] \gtreqless E[\Pi^{t*}(e_H)] \\ c_E \gtreqless p_H\left[\rho u^{-1}\left(\tfrac{k}{\rho}\right) - u^{-1}(k)\right] := \hat{c}_E\,. \end{gathered} \tag{4.42}$$

The behavior of $\hat{c}_E$ to changes in the parameters p_H, k, and ρ depends on the particular specification of the model. For a logarithmic specification of $u(x)$, one can show that the left hand side of the inequality strictly exceeds zero as long as $k > 1$. In this case, the amount of the evaluation cost is decisive for the choice of the evaluation frequency. Even if we take the logarithmic specification as given, it still depends on the chosen parameters whether risk aversion makes evaluations more or less desirable. The explanation for the effect of the stochastic shock being ambiguous is that, on the one hand, it lowers the value of an evaluation as with some positive probability no bonus will be paid, despite the evaluation having occurred and the agent having exerted effort. On the other hand, the damage for the agent is larger when he is not being paid due to the principal's bounded memory, as the bonus in the setup with the stochastic shock is larger.

4.5.2 Deviation Scenarios for the Principal

If the principal wants to cheat on the agent in a one-evaluation contract, two scenarios are imaginable. First, one could suppose that evaluations are contractible and, therefore, the principal cannot skip the scheduled evaluation at the end of the second period. Also, pretending to evaluate truthfully still comes at cost c_E. Then, the cheating consists in the fact that he forges the evaluation such that it does not plan any bonus for the agent's work in the two preceding periods, even though the agent has exerted high effort. So by deviating, he can save the money for the bonus, but not the evaluation cost. In a second scenario, evaluations are not contractible and the principal skips both, the evaluation itself and the payment of the bonus. For those scenarios, we will derive the critical interest rates r_1^o and r_2^o.

When we look at two-evaluation contracts, even more scenarios become possible. But, we think, the four we will identify here are quite exhaustive. All the deviations we consider have in common that the bonus corresponding to the second period will never be paid. In the first two scenarios, we presume that the principal will not pay any bonus. More precisely, the first scenario is analogous to the first scenario in the one-evaluation contract. As evaluations are contractible, the principal must conduct them, but he will not pay any bonus although he promised to do so. For a second scenario, one could imagine that evaluations are not contractible, but the agent will become suspicious if the principal skips the first evaluation. So he must at least evaluate and bear the corresponding cost after the first period, but afterwards, he will pay no bonus for either period and he does not evaluate a second time.

After all, we consider the following two scenarios to be more plausible. Still, the first evaluation is conducted, either because it is contractible or because the agent could otherwise learn that the principal is cheating. But the agent also learns the content of the evaluation and that information is contractible. So in a third scenario, the corresponding bonus becomes verifiable information after the first evaluation. Hence, if the principal cheated on the first-period bonus, the agent would learn it and, therefore, stop providing high effort. In this situation, the principal would be reluctant to cheat in the first period. When he deviates from the implicit contract, he has the entire revenue, but he pays only half the bonus and half the evaluation cost. A forth and last scenario we would like to tackle is similar to the second one under two evaluations, but also, once the first-period evaluation has been made, the agent learns the result and the bonus becomes contractible. We can find critical interest rates r_i^t, with $i = 1, 2, 3, 4$ that correspond to each of these scenarios. The following list contains the critical interest rates for one and two evaluations:

$$
\begin{aligned}
r_1^o &= \frac{2fe_H - e_H - \rho u^{-1}(\frac{e_H}{\rho}) - c_E}{(1+\rho)u^{-1}(\frac{e_H}{\rho})}, \\
r_2^o &= \frac{2fe_H - e_H - \rho u^{-1}(\frac{e_H}{\rho}) - c_E}{(1+\rho)u^{-1}(\frac{e_H}{\rho}) + c_E}, \\
r_1^t &= \frac{2fe_H - e_H - u^{-1}(e_H) - 2c_E}{2u^{-1}(e_H)}, \\
r_2^t &= \frac{2fe_H - e_H - u^{-1}(e_H) - 2c_E}{2u^{-1}(e_H) + c_E}, \\
r_3^t &= \frac{2fe_H - e_H - u^{-1}(e_H) - 2c_E}{u^{-1}(e_H)}, \\
r_4^t &= \frac{2fe_H - e_H - u^{-1}(e_H) - 2c_E}{u^{-1}(e_H) + c_E}.
\end{aligned}
$$

For both evaluation frequencies, the second threshold represents the strongest requirement to the relationship because the gains from a deviation are the highest is these scenarios. These are the scenarios discussed in Section 4.3.3.

Chapter 5

Competition for Workers

5.1 Introduction

In this chapter, we extend the model from Chapter 4 to consider differences in the principal's degree of forgetfulness. It seems unlikely that the parameter representing forgetfulness is equal for all employers. Therefore, we analyze a situation where heterogeneously forgetful principals compete for an agent. Competition between principals places the employee in the stronger position. While in a classical principal-agent model the employer makes a contract offer that sets the agent indifferent between signing it or not, the competitive market situation enables the agent to extract a positive utility.[1]

For two competing principals, the model is extended to a situation where one principal has a potential advantage due to a better memory capacity. The agent will choose to work for the principal who offers the most attractive contract. In a competitive setup with heterogeneous principals, we can show that the advantage we expected turns out to be relevant. The less forgetful principal is able to make

[1]Examples from the literature for competition between principals are Phelan (1995) (identical principals) and Manoli and Sannikov (2005) (heterogeneous principals).

a positive profit if one evaluation is chosen, and the more his competitor forgets, the larger this share of the surplus will be.

5.2 The Duopoly Setup

Having solved the basic model with a risk-averse agent, we will now extend the setup to a situation with two heterogeneously forgetful principals who compete for the labor of one agent.[2] The principals need the agent to produce output and the agent can only work for one of them, whilst the loser of the competition makes zero profit. In the following, principal j is more forgetful than principal i, i.e. $\rho_i > \rho_j$. To characterize the behavior of the competing employers, we must take a closer look at the options they have. Both principals will only make contract offers that create a positive expected surplus. Consequently, only when there exists at least one contract—with one or with two evaluations—for which the agent's productivity f is sufficiently high, the agent will be employed by one of the principals. As the critical profitability threshold belonging to a one-evaluation contract depends on the memory parameter ρ—exactly the parameter the principals differ in—we can state the following lemma on the corresponding thresholds:

Lemma 1 *The critical profitability threshold for a contract with one evaluation defined by (4.10) is strictly lower for principal i than the corresponding threshold for principal j, i.e. $f^o(\rho_j) := f^o_j > f^o(\rho_i) := f^o_i$.*

The proof is delegated to the appendix. The intuition of the preceding lemma is as follows: If a contract with one evaluation generates a positive surplus for

[2]The duopoly setup can be interpreted as a situation where some employers are able to generate signals on their ability to remember and can therefore build up heterogenous reputations of recruiting supervisors with strong memory capacities. Alternatively, the competition could take place between different departments within a firm. In this case it is plausible that employees have information on the memory capacities of supervisors in the departments.

principal i, this does not imply that the same is true for principal j. But the reverse implication—if a contract with one evaluation generates a positive surplus for principal j, the same holds for principal i—is true. Contracts with two evaluations generate identical surpluses for both principals. Hence, for an employment relationship to be established, at least one of the inequalities $f \geq f^t$ and $f \geq f_i^o$ must be true.

The chronology of the events in the two-principal setup is similar to the order of moves when there is only one employer. The main difference is that now not one but two employers simultaneously make a contract offer to the agent in Period 0. When they make their contract offers, the principals must consider their competitor's potential, i.e. profitable strategies and the criteria on which the agent will base his choice. To structure the following analysis, we will look at three important cases which are specified according to the available contract offers that create a positive surplus. We do not consider situations where only one principal can make an acceptable contract offer, i.e. we will only consider parameter constellations with competition between the employers.

The comparison of the thresholds shows us, which contract offer can realize the highest profits and who can, therefore, make the highest wage offer in competition. The lower the threshold, the higher the profit the corresponding contract yields. For the analysis, it is only important to look at contracts with thresholds that are below the agent's productivity f, because otherwise, the principal will make a loss when offering the contract. Furthermore, for each principal, only the threshold corresponding to the highest possible wage offer he can make is relevant. This allows us to restrict the analysis to three relevant competitive cases that are displayed in Table 5.2. An exhaustive characterization of possible cases can be found in the appendix.

Case	Productivity Thresholds	Evaluation(s)
A	$f \geq f_j^o \wedge f^t > f_j^o$	One
B	$f \geq f^t \wedge f_i^o < f^t$	Two
C	$f \geq f^t \wedge f_j^o > f^t > f_i^o$	One

Table 5.1: Overview of the Cases and their Characteristics

5.2.1 Case A: Only One-Evaluation Contracts

This case considers all situations where the most gainful contract either principal can offer is a contract with one evaluation. To derive the equilibrium outcome in case of competition between principals who offer one-evaluation contracts, we start by setting principal j's expected profit from an incentive-compatible contract equal to zero. To do so, we use $\beta^{o*} = \frac{1}{fe_H}u^{-1}(\frac{e_H}{\rho})$, as the entire amount of the wage payment can be adjusted via the participation constraint, while the incentive constraints in (4.2) and (4.3) continue to be fulfilled. Solving for α_i^o yields the highest fixed payment principal j can pay without making a loss:

$$E[\Pi_j^o(e_H)] \stackrel{!}{=} 0 \Leftrightarrow \alpha_j^o = 2fe_H - (1+\rho_j)u^{-1}\left(\frac{e_H}{\rho_j}\right) - c_E\,. \qquad (5.1)$$

Given this α_j^o and zero profit for principal j, the agent's expected utility when working for j is:

$$E[U_j^o(e_H|\Pi_j^o(e_H) = 0)] = 2fe_H - \rho_j u^{-1}\left(\frac{e_H}{\rho_j}\right) - e_H - c_E\,. \qquad (5.2)$$

His expected utility when employed by principal i is given by:

$$E[U_i^o(e_H)] = \alpha_i^o + u^{-1}\left(\frac{e_H}{\rho_i}\right) - e_H\,. \qquad (5.3)$$

To contract the agent, principal i must offer the agent an expected utility that is at least as large as the one he could possibly receive from principal j when the latter is making an expected profit of zero. So by setting (5.2) equal to (5.3) we can derive a critical fixed payment α_i^{o*} for which the employee is indifferent between working for i and working for j:

$$\alpha_i^{o*} = 2fe_H - u^{-1}\left(\frac{e_H}{\rho_i}\right) - \rho_j u^{-1}\left(\frac{e_H}{\rho_j}\right) - c_E\,. \tag{5.4}$$

If principal i sets α_i^{o*} according to (5.4), the agent is ready to work for him and his profit is:

$$E[\Pi_i^{o*}(e_H)] = \rho_j u^{-1}\left(\frac{e_H}{\rho_j}\right) - \rho_i u^{-1}\left(\frac{e_H}{\rho_i}\right)\,. \tag{5.5}$$

Of course, we are interested whether this profit can be positive and when this is the case. Proposition 6 states that the less forgetful principal can actually make a strictly positive profit:

Proposition 6 *Given that there is competition and contracts with one evaluation yield the highest surplus, the following holds: The agent accepts working for principal i under a one-evaluation contract that makes him indifferent between working for one or the other principal ($E[U_i^o(e_H)] = E[U_j^o(e_H)]$) and which leaves principal j with a profit of zero ($E[\Pi_j^o(e_H)] = 0$). Principal i, the principal with the better memory, expects to realize a strictly positive profit ($\Pi_i^o(e_H) > 0$).*

Proof. The profit in (5.5) already takes principal j's zero-profit condition and the indifference of the agent into account. So we only need to prove that $\Pi_i^o(e_H) > 0$ which can be reformulated as $\rho_j u^{-1}(\frac{e_H}{\rho_j}) > \rho_i u^{-1}(\frac{e_H}{\rho_i})$. As this is the same inequality as in (5.13) the further proof is identical to the one of Lemma 1. ■

When the offer complies with (5.4), the agent's expected utility from working for principal i is:

$$E[U_i^o(e_H)] = 2fe_H - \rho_j u^{-1}\left(\frac{e_H}{\rho_j}\right) - e_H - c_E\,. \tag{5.6}$$

Note that his utility only depends on ρ_j but is independent of his employer's degree of forgetfulness ρ_i. For the surplus, we obtain the opposite result:

$$E[S_i^o(e_H)] = 2fe_H - \rho_i u^{-1}\left(\frac{e_H}{\rho_i}\right) - e_H - c_E\,. \tag{5.7}$$

Thus, the surplus is a function of the actual employer's memory parameter only. This is quite logical, as ρ_i induces the risk the agent has to bear despite his risk aversion. On the other hand, ρ_j leaves the surplus unaffected, but it determines the actual split which employer i can negotiate, and this again depends on the agent's outside option with principal j as employer. The surplus is identical to the principal's profit from equation (4.6) in the non-competitive setup.

The agent's expected compensation is:

$$E[C^o(e_H)] = 2fe_H + \rho_i u^{-1}\left(\frac{e_H}{\rho_i}\right) - \rho_j u^{-1}\left(\frac{e_H}{\rho_j}\right) - c_E\,. \tag{5.8}$$

As in (6.9), the monetary value the expected compensation has for the principal exceeds the expected disutility of effort of the agent.

5.2.2 Case B: Only Two-Evaluation Contracts

In this case, a contract with two evaluations offered by either principal generates the highest surplus. It is possible, that profitability is also given for contracts with one evaluation, but such a contract will never be offered as it can always be overbid by a contract with two evaluations. If only two-evaluation contracts play a role, the surplus at stake is independent of ρ and the two principals can offer the same range of profitable two-evaluation contracts. As the principals compete in the amount of the wage payment for a given level of effort, it is reasonable to apply a Bertrand argumentation here. If one principal makes a contract offer with an expected wage payment w that leaves him in expectations with a positive profit, the other principal can still offer a slightly higher wage

$w + \epsilon$ and contract the agent. This is true until the chosen wage scheme induces an expected profit of zero. As both employers anticipate this overbidding à la Bertrand, they offer contracts with the same wage scheme which makes them indifferent between employing the agent and not producing. Consequently, the agent who faces such contracts is indifferent which of the two he will sign.

Due to the perfect preservation of the performance information in an employment relationship with two evaluations, forgetfulness plays no role in the following calculus. The principals are symmetric again, and we can skip the indexes i and j for the compensation variables. Using the incentive-compatible bonus factor $\beta^{t*} = \frac{1}{fe_H} u^{-1}(e_H)$ from Section 4.2.3 we can derive the fixed payment α^{t*} that leads to zero expected profit for the contracting principal:

$$E[\Pi^t(e_H)] \stackrel{!}{=} 0 \Leftrightarrow \alpha^{t*} = 2[fe_H - u^{-1}(e_H) - c_E]\,. \tag{5.9}$$

Given these compensation parameters, the agent's expected utility for two evaluations is:

$$\begin{aligned} E[U^{t*}(e_H)] &= \alpha^{t*} + u(\beta^{t*} fe_H) + \beta^{t*} fe_H - 2e_H \\ &= 2fe_H - e_H - u^{-1}(e_H) - 2c_E\,. \end{aligned} \tag{5.10}$$

This corresponds to the principal's expected profit in Section 4.2.3.

5.2.3 Case C: One Evaluation by Principal i, Two by j

This last case considers a situation where the contract yielding the highest surplus is a one-evaluation contract with principal i. At the same time, principal j can offer a profitable contract with two evaluations. Hence, the agent has an outside option when negotiating with principal i. We can take the expected utility the agent gets when working for principal j from equation (5.10) in Case B. Now, the wage offer by principal i for a contract with one evaluation must at least give the agent the same expected utility:

$$\begin{aligned} & E[U_i^o(e_H)] = E[U^t(e_H|\Pi^t(e_H) = 0)] \\ & \Leftrightarrow \alpha_i^{o*} = (2f-1)e_H - u^{-1}\left(\tfrac{e_H}{\rho_i}\right) - 2c_E\,. \end{aligned} \tag{5.11}$$

Principal i's profit with such a contract is:

$$E[\Pi^{o*}(e_H)] = e_H - \rho u^{-1}\left(\frac{e_H}{\rho}\right) + c_E\,. \tag{5.12}$$

Again, the surplus, i.e. the sum of (5.10) and (5.12), is identical to the principal's profit from equation (4.6) in the non-competitive setup.

5.2.4 Results from the Duopoly Setup

By analyzing the cases A to C, we have characterized the interplay of two principals who differ in their memory capacity. They are the only cases featuring a strictly competitive situation. Our findings in these cases are quite intuitive. The principal who has a better memory can only benefit from his advantage when a contract with one evaluation is efficient. In this case, the surplus is divided between this principal and the agent. When it is efficient to evaluate twice, the entire surplus goes to the agent, as the principals are homogeneous in their offers. Compared to Section 4.2.3, where the principal has the entire market power, the surplus is unchanged. Only its distribution varies when market power is shifted to the agent.

In this formal analysis, we have considered a risk-averse agent. Intuitively, the outcome of the model with a risk-neutral agent is quite straightforward. The agent has the entire market power in the duopoly setup with two principals, as the latter will overbid their respective compensation offers until both make zero profits. With risk neutrality, the surplus at stake is independent of the degree of forgetfulness. Therefore, the principals' heterogeneity offers no advantage to the less forgetful one, i.e. if principal i gets zero profit from an offer, the same offer implies zero profit for principal j and the other way around. Although, we have assumed an asymmetry in forgetfulness the outcome is symmetric. The agent is indifferent between working for principal i or working for j and receives the entire surplus.

5.3 Conclusion

The extension to a duopoly setup under risk aversion does not change the decision on the optimal evaluation frequency. In cases A and C, the principal who has the better memory capacity has an advantage over his competitor. When the agent is only offered a contract with one evaluation, he decides to work for the less forgetful principal and the latter can realize a strictly positive profit. The more the competitor forgets, the larger this share of the surplus will be. Since the distribution of market power does not affect the evaluation frequency chosen, efficiency is preserved in the duopoly framework.

5.4 Appendix

5.4.1 Proof of Lemma 1

Proof. The inequality in the lemma can be reformulated as:

$$\rho_j u^{-1}\left(\frac{e_H}{\rho_j}\right) > \rho_i u^{-1}\left(\frac{e_H}{\rho_i}\right). \tag{5.13}$$

To show that this holds, we prove that the derivative of $\rho u^{-1}(\frac{e_H}{\rho})$ with respect to ρ is negative. This is sufficient as $\frac{\partial(\rho u^{-1}(\frac{e_H}{\rho}))}{\partial\rho} < 0$ implies that $\rho_j u^{-1}(\frac{e_H}{\rho_j})$ must be larger than $\rho_i u^{-1}(\frac{e_H}{\rho_i})$ and, therefore, $\Pi_i^o(e_H) > 0$ is true. The derivative is:

$$\begin{aligned}\frac{\partial(u^{-1}(\frac{e_H}{\rho}))}{\partial\rho} &= u^{-1}(\tfrac{e_H}{\rho}) + \rho u_\rho^{-1}(\tfrac{e_H}{\rho}) \cdot (-\tfrac{e_H}{\rho^2}) \\ &= u^{-1}(\tfrac{e_H}{\rho}) - \tfrac{e_H}{\rho} \cdot u_\rho^{-1}(\tfrac{e_H}{\rho}),\end{aligned} \tag{5.14}$$

where $u_\rho^{-1}(\cdot)$ stands for the first derivative of the inverse utility function with respect to ρ. Now the condition $\frac{\partial(\rho u^{-1}(\frac{e_H}{\rho}))}{\partial\rho} < 0$ can be written as:

$$\begin{aligned}\frac{\partial(\rho u^{-1}(\frac{e_H}{\rho}))}{\partial\rho} < 0 \;&\Leftrightarrow u^{-1}(\tfrac{e_H}{\rho}) - \tfrac{e_H}{\rho} \cdot u_\rho^{-1}(\tfrac{e_H}{\rho}) < 0 \\ &\Leftrightarrow \frac{u^{-1}(\frac{e_H}{\rho})}{\frac{e_H}{\rho}} < u_\rho^{-1}(\tfrac{e_H}{\rho}).\end{aligned} \tag{5.15}$$

Since it is the inverse of the concave function $u(\cdot)$, the function $u^{-1}(\cdot)$ is convex. Therefore, the inequality in (5.15) holds. ■

5.4.2 Division into Cases

Case 1: $f < f^t$: This case contains three subcases of which only the last one is really interesting.

- Case 1.1: $f_j^o > f_i^o > f$
- Case 1.2: $f_j^o > f \geq f_i^o$
- Case 1.3: $f \geq f_j^o > f_i^o$

Case 2: $f \geq f^t$: Contains all situations where two evaluations are profitable and it is subdivided into three groups of subcases:

- Case 2.1: $f_i^o > f^t$
  - Case 2.1.1: $f_j^o > f_i^o > f \geq f^t$
  - Case 2.1.2: $f_j^o > f \geq f_i^o > f^t$
  - Case 2.1.3: $f \geq f_j^o > f_i^o > f^t$

- Case 2.2: $f_j^o > f^t > f_i^o$
  - Case 2.2.1: $f_j^o > f \geq f^t > f_i^o$
  - Case 2.2.2: $f \geq f_j^o > f^t > f_i^o$
- Case 2.3: $f \geq f^t > f_j^o > f_i^o$

The comparison of the thresholds shows us which contract offer can realize the highest profits and, therefore, who can make the highest wage offer in a competitive situation. The lower the threshold, the higher the profit the corresponding contract yields. For the analysis, it is only important to look at contracts with thresholds that are below the agent's productivity f, because otherwise, the principal would make a loss when offering the contract. Furthermore, for each principal only the threshold corresponding to the highest possible wage offer he can make is relevant. This allows us to reduce the analysis to three different cases:

- Case A: $f \geq f_j^o \wedge f^t > f_j^o$
- Case B: $f \geq f^t \wedge f_i^o > f^t$
- Case C: $f \geq f^t \wedge f_j^o > f^t > f_i^o$

Case A is the situation where the two contracts with one evaluation offered by the principals are the most gainful proposals available. This case contains the cases 1.3—the only situation that can be labeled *competitive* in Case 1, where it is profitable for both principals to offer a contract with one evaluation—and 2.3. The first and the second subcase of Case 1 are of minor interest. In Case 1.1, none of the principals can offer a gainful contract so the agent will be employed by neither. In the second subcase, only principal i can offer a contract the agent might accept. Only one gainful contract exists: A contract proposed by principal i which contains one evaluation. Principal i will offer a contract that guarantees the agent an expected utility equal to zero. The analysis is therefore identical to what has been treated in Section 4.2.2. Since the agent's only outside option is refusing to work, the employer can appropriate the entire surplus.

Case 2.1 and all its subcases refer to situations where two evaluations are always more profitable than one. Therefore the principals compete offering two-evaluation contracts. These cases are resumed in Case B. The third important case, Case C, contains cases where the contract offering the highest surplus is a one-evaluation contract by principal i. The second-best option is a contract offer with two evaluations. This is true for all cases belonging to 2.2.

Chapter 6

Heterogenous Agents

6.1 Introduction

In this chapter, we investigate the effect of heterogeneity of workers on optimal evaluation frequency. In particular, we want to find out whether employers should evaluate all of their employees at the same frequency or whether it can be optimal to offer contracts with differing evaluation frequencies at the same time. We consider a situation with two types of employees: For one of them it is more costly to produce a given level of output than for the other type of employee. So, potential employees indirectly differ in their productivity. This characteristic is not observable for the employer who wants to hire an agent. It is private information to the employee. Thus, the employer could offer contracts with different evaluation frequencies to screen the agents despite the information asymmetry. This is possible as the frequency of evaluations is directly linked to incentives because it enhances the accuracy of the outcome of the evaluation and, hence, reduces the risk attached to possible bonus payments.

Again, we motivate the increase of accuracy through more evaluations by a memory bias on behalf of the supervisor. The longer the period between the creation of output and the judgment of its value, the less likely the supervisor is to remember the real value of the output correctly. So the payoff of more evaluations comes through their impact on incentives. The more precisely an evaluation reflects the actual performance of an agent, the higher the probability that a high output will be valued through a bonus and, consequently, the higher the incentives to deliver such output. Of course, the results of the model apply to any situation where the quality of the evaluation outcome increases in evaluation frequency.

More frequent evaluations should not only lead to more precise evaluation outcomes, but they are very likely to be more costly at the same time. So, in a first step, the employer must trade off the cost of additional evaluations against their indirect payoff through stronger incentives. Furthermore, the existence of asymmetric information might encourage him to adjust the evaluation frequency to an optimal screening policy. Agents with high productivity might prefer to be evaluated more often as the loss they suffer due to a less accurate evaluation is larger than the one of a less productive agent. In contrast, low-productivity agents loose more, when choosing a high evaluation frequency as the indirect evaluation cost they incur through a lower fixed wage pays off relatively less for them.

We capture the principal's decision problem in a model inspired by Moen and Rosén (2005). The authors analyze the screening of heterogeneously productive agents who are evaluated on the basis of a distorted performance measure, i.e. a measure reacting differently to two separate tasks. In turn, our analysis proceeds analogously, but it refers to two different periods rather than tasks. Due to the employer's memory bias, the measured performance is the less distorted the more often an agent is evaluated. So, in terms of Moen and Rosén (2005), more frequent evaluations soften the underlying distortion. Of course, this comes at a cost, as total evaluation cost increases with an additional evaluation. The present model investigates in particular, which contracts an employer will offer when he is incompletely informed about the agents' productivity. He only knows the distribution of types but cannot tell the type of an agent he faces when making a contract offer.

We find that an employer will never offer contracts with different evaluation frequencies when information about the agents' type is symmetric. In this case, the choice of the optimal evaluation frequency is independent of the type and all agents should be evaluated at the same rate. When we come to the analysis of a situation in which information about his type is private to the agent, this result is only weakened under some restrictions. In the general model, screening will only be realized by the use of the compensation parameters. This is the case as the optimal evaluation frequency is independent of the type. When the principal has two parameters at hand to differentiate contracts, namely the compensation parameters, this will be sufficient for optimal screening. Screening contracts that differentiate with respect to evaluation frequency are always outperformed by one of the screening contracts with equal evaluation frequency. Also, pooling heterogenous agents will never be an equilibrium.

Only when contract design is subject to restrictions, such as the obligation to compensate employee performance in an equal way, screening by the use of different evaluation frequencies can be worthwhile. In this case, the use of evaluation frequencies as a screening device can become an equilibrium strategy for an uninformed principal. If the variation of bonus factors between groups of employees is precluded, equal evaluation frequencies automatically imply pooling of the agents. Depending on the distribution of the respective types in the population and other exogenous factors, both pooling contracts as well as both contracts that screen via evaluation frequency can become an equilibrium offer for the principal.

In Section 6.2, we extend the basic model of Moen and Rosén (2005) to a setup where a forgetful monopsonistic employer can decide on the evaluation frequency for his employees. We extend the analysis to a model with heterogeneous agents in

Section 6.3 and investigate possible contract offers with asymmetric information in Section 6.4. Section 6.5 concludes this chapter.

6.2 The Model with Homogenous Workers

The model we propose in this section is strongly related to a framework proposed by Moen and Rosén (2005). In a labor market, a principal contracts with risk-neutral agents to produce an output. Note that contrary to Moen and Rosén (2005), we assume the principal to have monopsony power. First, we look at homogenous workers who all have the same cost of effort provision. Second, we extend the model to heterogenous workers in Sections 6.3 and 6.4. The principal offers two-period contracts to workers who produce a stochastic output that is worth

$$\pi = e_1 + e_2 + \varepsilon \,, \tag{6.1}$$

where e_1 and e_2 is the level of effort in period $t = 1, 2$, and ε is a symmetric random shock on output with mean zero.[1] As the principal has bounded memory, the value of output he observes is

$$\tilde{\pi} = \rho \, e_1 + e_2 + \varepsilon := \tilde{e} + \varepsilon \,. \tag{6.2}$$

The factor ρ represents the probability of the principal to remember an agent's output. It is updated to one, if the principal makes an evaluation at the end of the first period, otherwise ρ lies somewhere between zero and one, i.e. $\rho \in \,]0, 1[$. Agents face a convex effort cost which is additively separable in periods:

$$C(e_1, e_2) = \frac{(e_1 - e^0)^2}{2} + \frac{(e_2 - e^0)^2}{2} \,, \tag{6.3}$$

[1]For a more precise interpretation, ε is the sum of two identically distributed shocks occurring in each period. There are no stochastic dependencies between periods.

for $e_1, e_2 \geq e^0$. An agent's utility is composed of his expected wage payment minus the effort cost:

$$E[U(e_1, e_2)] = E[w(\tilde{e})] - C(e_1, e_2)\,. \tag{6.4}$$

As the wage payment will be based upon the observed aggregate output $\tilde{\pi}$, the agent will choose an effort combination (e_1, e_2) such that he minimizes his overall effort cost for a given aggregate effort level $\tilde{e}$. In expectations, this corresponds to the following minimization problem:

$$\min_{e_1, e_2} \left[\frac{(e_1 - e^0)^2}{2} + \frac{(e_2 - e^0)^2}{2} \right] \quad s.\,t. \;\; \rho\, e_1 + e_2 = \tilde{e}\,, \tag{6.5}$$

where $\tilde{e}$ is also the expected output to be observed. The solution to the problem can be characterized by the equation

$$\frac{e_2 - e^0}{e_1 - e^0} = \frac{1}{\rho}\,. \tag{6.6}$$

So independent of the observed level of output, the worker will optimally choose to exert more effort in period two than in period one as long as ρ is not equal to one, i.e. $e_1 < e_2$ for $\rho \in \,]0, 1[$. Following the approach of Moen and Rosén (2005), the firm's expected output can be rewritten as

$$e_1 + e_2 = \frac{1 + \rho}{1 + \rho^2}\,\tilde{e} + \frac{(1 - \rho)^2}{1 + \rho^2}\,e^0\,. \tag{6.7}$$

The derivative of equation (6.7) with respect to $\tilde{e}$ is

$$\frac{\partial(e_1 + e_2)}{\partial \tilde{e}} = \frac{1 + \rho}{1 + \rho^2} = \beta^*\,. \tag{6.8}$$

Hence, the principal should pay the agent a bonus of β^* for each unit of aggregate output $\tilde{e}$ he expects to observe, as this corresponds to the firm's expected marginal gain from measured output. We concentrate on linear compensation schemes, where the agent receives an expected wage payment that is composed of a fixed wage α and a bonus calculated on the basis of the product of β^* and the observed output level $\tilde{e}$. Therefore, the expected wage payment for both periods is

$$E[w(\tilde{e})] = \alpha + \beta^* \, \tilde{e} \, . \tag{6.9}$$

Knowing the expected output from equation (6.7), we can now derive the principal's expected profit:

$$E[\Pi] = e_1 + e_2 - \alpha - \beta^* \, \tilde{e} - c = \frac{(1-\rho)^2}{1+\rho^2} e^0 - \alpha - c \, , \tag{6.10}$$

where $c \in \{c_o, c_t\}$, with $c_o < c_t$, is the evaluation cost referring to a low or a high evaluation frequency respectively, i.e. one or two evaluations in a two-period model. Consequently, if $c = c_o$ then $\rho < 1$ and if $c = c_t$ then an updating of ρ occurs and leads to $\rho = 1$. To derive the agent's equilibrium effort level, we first calculate the agent's effort cost for the provision of an aggregate effort $\tilde{e}$.[2]

$$C(\tilde{e}) = \frac{(\tilde{e} - (1+\rho)e^0)^2}{2(1+\rho^2)} \, . \tag{6.11}$$

Hence, the expected utility for a given effort level $\tilde{e}$ equals

$$E[U(\tilde{e})] = \alpha + \beta^* \, \tilde{e} - C(\tilde{e}) = \alpha + \beta^* \, \tilde{e} - \frac{(\tilde{e} - (1+\rho)e^0)^2}{2(1+\rho^2)} \, . \tag{6.12}$$

The agents maximize their utility with respect to $\tilde{e}$:

$$\frac{\partial E[U(\tilde{e})]}{\partial \tilde{e}} \overset{!}{=} 0 \Leftrightarrow \tilde{e}^* = (1+\rho^2)\beta^* + (1+\rho)e^0 \Leftrightarrow \tilde{e}^* = (1+\rho)(1+e^0) \, . \tag{6.13}$$

Therefore, the expected output in equilibrium simplifies to

$$e_1 + e_2 = 2e^0 + \frac{(1+\rho)^2}{1+\rho^2} = 2e^0 + (1+\rho)\beta \, . \tag{6.14}$$

[2] A detailed derivation of this cost function can be found in the appendix to this chapter.

Given the optimal level of aggregate effort, an agent's expected utility is characterized by

$$E[U(\tilde{e}^*)] = \alpha + \frac{1}{2}\beta^{*2}(1+\rho^2) + \beta^*(1+\rho)e^0 = \alpha + \frac{(1+\rho)^2}{1+\rho^2}\left(e^0 + \frac{1}{2}\right). \quad (6.15)$$

As the principal is a monopsonist, he will choose α such that $U(\tilde{e}^*)$ equals an agent's outside option $\bar{U}$ in equilibrium:

$$\alpha^* = \bar{U} - \frac{(1+\rho)^2}{1+\rho^2}\left(e^0 + \frac{1}{2}\right). \quad (6.16)$$

The expected equilibrium profit is given by

$$E[\Pi] = \frac{(1-\rho)^2}{1+\rho^2}e^0 - \alpha^* - c\,. \quad (6.17)$$

Substituting (6.16) into (6.17) leads us to an equilibrium expected profit of the principal which amounts to

$$E[\Pi^{o*}] = 2e^0 + \frac{(1+\rho)^2}{2(1+\rho^2)} - \bar{U} - c \quad (6.18)$$

for a situation where the principal faces no competitors. To make inferences on the principal's choice concerning an optimal evaluation frequency in the two-period contract, we compare his profit for an optimal one-evaluation contract ($E[\Pi^{o*}(\rho < 1, c_o)]$) with the profit an optimal two-evaluation contract ($E[\Pi^{t*}(\rho = 1, c_t)]$):[3]

$$E[\Pi(\rho < 1, c_o)] \gtreqless E[\Pi(\rho = 1, c_t)] \Leftrightarrow \Delta c \gtreqless \frac{(1-\rho)^2}{2(1+\rho^2)}\,, \quad (6.19)$$

where $\Delta c = c_t - c_o$ is the additional cost from a second evaluation.

[3]The reader should remember here, that a switch to two evaluations leads to a higher evaluation cost c_t but at the same time it updates ρ to one.

Proposition 7 *If a second evaluation causes an additional cost, the principal faces a trade-off between offering a one- and offering a two-evaluation contract to the risk-neutral agent. This trade-off is independent of the agent's cost parameter* e^0.

Proof. The right-hand side of the inequality in (6.19) is strictly larger than zero and the left-hand side does not depend on the cost parameter e^0. ■

So, as soon as a second evaluation causes an additional cost, the principal faces a trade-off between this cost and the disposal over more efficient incentives due to the information that is gained by another evaluation. The result that the principal faces a trade-off concerning the evaluation frequency—independent of the agent's cost parameter—is complementary to the finding from Chapter 4 where risk-aversion suffices for the existence of a trade-off. Hence, not only when facing risk-averse agents, the choice of evaluation frequency highly depends on the particular situation the employer faces.

6.3 Heterogeneity and Symmetric Information

In this section, we will analyze how heterogeneity with respect to the agents' effort cost impacts the model. In Section 6.2, we assumed that all agents face the same function (6.3) determining the cost of the provision of a given output level e_1 or e_2. The higher the parameter e^0 in this function, the lower the agent's effort cost. To model heterogeneity, we allow e^0 to take two values, i.e. $e^0 \in \{e_L, e_H\}$ with $e_L < e_H$. This gives us two types of agents, an L-agent who faces a higher effort cost and an H-agent whose effort cost for a given output level is strictly lower.

Before we consider asymmetric information in Section 6.4, we will interpret the results from Section 6.2 for heterogenous agents whose types are observable for the employer. Since the principal can optimally design the contracts offered, as derived in the previous section, this is considered a benchmark case. Equation (6.13) gives us the intuitive result that an H-agent will produce more than an L-agent, due to his lower effort cost. Therefore, the output in equation (6.14) increases in e^0. This is also true for the principal's profit in equation (6.18). The optimal bonus parameter β^* is invariant to the cost parameter e^0, and we can see from equation (6.16) that agents with lower effort cost need a lower fixed wage to participate in the contract. As already mentioned, we can deduce from the inequality in (6.19) that the principal's optimal decision on the evaluation frequency is independent of the agent's type as long as the principal can observe the latter to be able to set compensation parameters accordingly. This is an important result, as it suggests that a differentiation of contracts through the evaluation frequency—if it occurs—must be caused by asymmetric information.

6.4 Heterogeneity and Asymmetric Information

In the following, we will investigate how the optimal contract parameters and particularly the optimal evaluation frequency are affected when information about the type is private to the agent. So, while the agent is fully informed, the principal only knows the distribution of types within the population of agents he faces. This population consists of a fraction a representing L-type agents and a fraction $(1-a)$ of H-type agents. Put differently, a is the probability of a randomly drawn agent to be of the L-type and $(1-a)$ the respective probability for an H-type. In this situation, the principal might be interested in screening the agents by proposing two kinds of contracts such that each worker type selects himself into a specific contract.

To screen the agents, the principal theoretically can use three instruments to differentiate contracts. He can offer contracts that differ concerning the fixed wages, the bonus payments, and the evaluation frequency. It is easy to understand why differentiating only with respect to one compensation parameter or to the evaluation frequency, keeping the other contract characteristics identical for all agents, is useless for screening. Independent of their type, employees strictly prefer a higher bonus factor, a higher fixed wage, and a higher evaluation frequency to lower ones. Therefore, a contract that successfully screens the agents must differ in two respects at least.

But should the screening contracts differ in more than two respects? We find that in an unrestricted model (Section 6.4.1), it is sufficient to use the compensation parameters to screen the agents under asymmetric information. The optimal evaluation frequency is unaffected and identical for both types of agents (this is a logical consequence of the independence result in inequality (6.19)). Only if we restrict the principal's possibility to design contracts (Section 6.4.2), e.g. if he cannot accord different bonuses to employees who have achieved the same score in an evaluation, the choice of the evaluation frequency will be used for screening. We will now characterize the different linear pooling and separating contracts a principal can offer to the agents.[4] Having derived the optimal contract parameters for each type of contract, we can compare their profitability in a next step.

6.4.1 The Unrestricted Model

Pooling Contracts

As he is a monopsonist, one of the principal's options is to pool the agents using a single contract with one or two evaluations. We will first solve for the optimal one-evaluation contract $(\alpha^{oP}, \beta^{oP})$ that attracts both types of workers. In the following section, the corresponding results for the two-evaluation setup are given.

[4]Additionally, one could also extend analysis to the possibility of shutting down the L-type. Here, we chose to focus on contracts which lead to employment of both types of agents. Already, this leaves us with four relevant cases to compare.

One Evaluation

An optimal pooling contract with only one evaluation solves the following problem:

$$\max_{\alpha^{oP},\beta^{oP}} aE[\Pi^{oP}(e_L)] + (1-a)E[\Pi^{oP}(e_H)] \quad s.t. \tag{6.20}$$

$$E[U^H(\tilde{e}^{oP*},\alpha^{oP},\beta^{oP})] \geq 0 \tag{6.21}$$

$$E[U^L(\tilde{e}^{oP*},\alpha^{oP},\beta^{oP})] \geq 0\,. \tag{6.22}$$

The principal maximizes his expected utility subject to the participation constraint of each type of agent and given optimal effort allocation on behalf of the agents as derived in (6.13). From (6.15), we know the agent's maximized utility for a given β. Therefore, we deduce that the conditions in (6.21) and (6.22) equal

$$E[U^H(\tilde{e}^{oP*},\alpha^{oP},\beta^{oP})] = \alpha^{oP} + \frac{1}{2}(\beta^{oP})^2(1+\rho^2) + \beta^{oP}(1+\rho)e_H \geq 0 \tag{6.23}$$

and

$$E[U^L(\tilde{e}^{oP*},\alpha^{oP},\beta^{oP})] = \alpha^{oP} + \frac{1}{2}(\beta^{oP})^2(1+\rho^2) + \beta^{oP}(1+\rho)e_L \geq 0 \tag{6.24}$$

respectively. Subtracting (6.24) from (6.23) yields:

$$\beta^{oP}(1+\rho)[e_H - e_L] \geq 0\,. \tag{6.25}$$

Hence, the H-agent's participation constraint in (6.23) is always slack when the L-agent's constraint in (6.24) is fulfilled with equality. To formulate the objective function in (6.20), we need the principal's expected profit from a single one-evaluation contract given the agent's type and optimal effort choice:

$$\begin{aligned} &E[\Pi^{oP}(\tilde{e}^{oP*},\alpha^{oP},\beta^{oP})] = \\ &\left[2-(1+\rho)\beta^{oP}\right]e^0 + (1+\rho)\beta^{oP} + (1+\rho^2)(\beta^{oP})^2 - \alpha^{oP} - c_o\,. \end{aligned} \tag{6.26}$$

Consequently, the principal's aggregate expected profit over the entire population of agents is given by:

$$\begin{aligned}
&E[\Pi^{oP}(a)] = aE[\Pi^{oP}(e_L)] + (1-a)E[\Pi^{oP}(e_H)] = \\
&\left[2 - \beta^{oP}(1+\rho)\right]\left[ae_H + (1-a)e_L\right] \\
&+(1+\rho)\beta^{oP} - (\beta^{oP})^2(1+\rho^2) - \alpha^{oP} - c_o\,.
\end{aligned} \tag{6.27}$$

The Lagrangian representing the principal's maximization problem is

$$\mathcal{L} = E[\Pi^{oP}(a)] - \lambda\left[\alpha^{oP} + \frac{1}{2}(\beta^{oP})^2(1+\rho^2) + \beta^{oP}(1+\rho)e_L\right]\,. \tag{6.28}$$

Differentiating (6.28) with respect to α^{oP}, β^{oP}, and λ yields

$$\frac{\partial\mathcal{L}}{\partial\alpha^{oP}} = -1 - \lambda \overset{!}{=} 0 \Leftrightarrow \lambda = -1\,, \tag{6.29}$$

$$\begin{aligned}
&\tfrac{\partial\mathcal{L}}{\partial\beta^{oP}} = -(1+\rho)[ae_L + (1-a)e_H - 1] - 2\beta^{oP}(1+\rho^2) \\
&-\lambda[\beta^{oP}(1+\rho^2) + (1+\rho)e_L] \overset{!}{=} 0\,,
\end{aligned} \tag{6.30}$$

and

$$\frac{\partial\mathcal{L}}{\partial\lambda} = \alpha^{oP} + \frac{1}{2}(\beta^{oP})^2(1+\rho^2) + \beta^{oP}(1+\rho)e_L \overset{!}{=} 0\,. \tag{6.31}$$

Substituting (6.29) into (6.30) gives us the optimal bonus factor

$$\beta^{oP*} = \frac{1+\rho}{1+\rho^2}q\,, \tag{6.32}$$

where $q = 1 - (1-a)(e_H - e_L)$. This expression allows to solve for the optimal fixed wage

$$\alpha^{oP*} = -\frac{1}{2}(\beta^{oP})^2(1+\rho^2) - \beta^{oP}(1+\rho)e_L = -\frac{(1+\rho)^2}{2(1+\rho^2)}q[q+2e_L]\,. \tag{6.33}$$

Given the optimal compensation parameters for a pooling contract with one evaluation, the H-type agent realizes an expected utility of

$$E[U^H(\tilde{e}^{oP*}, \alpha^{oP*}, \beta^{oP*})] = \frac{(1+\rho)^2}{1+\rho^2} q(e_H - e_L)\,. \tag{6.34}$$

Hence, the H-agent's utility increases in ρ. The utility is hump-shaped in $(e_H - e_L)$ with a maximum at $(e_H - e_L) = \frac{1}{2(1-a)}$. In turn, the expected utility of the L-agent equals zero in equilibrium. Finally, the principal's expected profit amounts to

$$E[\Pi^{oP*}(a)] = 2[ae_L + (1-a)e_H] + \frac{(1+\rho)^2}{2(1+\rho^2)} q^2 - c_o\,. \tag{6.35}$$

Two Evaluations

The solution for a pooling contract with two evaluations can be derived using the results of the previous sections. One simply has to update $\rho = 1$ and replace the lower value of the evaluation cost c_o by the higher evaluation cost c_t. The optimal compensation parameters are

$$\beta^{tP*} = 1 - (1-a)(e_H - e_L) = q \tag{6.36}$$

and

$$\alpha^{tP*} = -q(q + 2e_L)\,. \tag{6.37}$$

Given the two-evaluation contract, the H-agent can again extract a positive expected utility which equals

$$E[U^H(\tilde{e}^{tP*}, \alpha^{tP*}, \beta^{tP*})] = 2q(e_H - e_L)\,, \tag{6.38}$$

the L-agent's expected utility is zero, and the principal's expected profit equals

$$E[\Pi^{tP*}(a)] = 2[ae_L + (1-a)e_H] + q^2 - c_t\,. \tag{6.39}$$

Separating Contracts with Equal Evaluation Frequencies

Instead of proposing only one type of contract, the employer can address the heterogeneity of types by offering different contracts at the same time. More particularly, to be able to screen the agents he can offer a menu of contracts with each contract consisting of a different vector of contract parameters, possibly implying different evaluation frequencies. As already mentioned, differentiating only with respect to one parameter or to the evaluation frequency, keeping the other contract characteristics identical for all agents, induces no screening. In this section, we investigate contracts that screen via the differentiation of compensation variables only. In Section 6.4.2, the screening activity of the employer will additionally allow for different evaluation frequencies within the menu of contracts.

One Evaluation

In a separating contract, where both agents are evaluated once, the principal's expected profit is

$$\begin{aligned} E[\Pi^{oo}(a)] = a[(2-(1+\rho)\beta_L^{oo})e_L + (1+\rho)\beta_L^{oo} - (1+\rho^2)(\beta_L^{oo})^2 - \alpha_L^{oo} - c_o] \\ +(1-a)[(2-(1+\rho)\beta_H^{oo})e_H + (1+\rho)\beta_H^{oo} - (1+\rho^2)(\beta_H^{oo})^2 - \alpha^{oo_H} - c_o] \ . \end{aligned} \tag{6.40}$$

He maximizes (6.40) with respect to the H-agent's incentive and the L-agent's participation constraint.

$$\max_{\alpha_L^{oo},\beta_L^{oo},\alpha_H^{oo},\beta_H^{oo}} \quad E[\Pi^{oo}(a)] \quad s.t. \tag{6.41}$$

$$E[U^H(\tilde{e}^{oo*},\alpha_H^{oo},\beta_H^{oo})] \quad \geq \quad E[U^H(\tilde{e}^{oo*},\alpha_L^{oo},\beta_L^{oo})] \tag{6.42}$$

$$E[U^L(\tilde{e}^{oo*},\alpha_L^{oo},\beta_L^{oo})] \quad \geq \quad 0\,. \tag{6.43}$$

For a contract to be accepted by L-agents, the optimal compensation parameters are

$$\alpha_L^{oo*} = -\frac{(1+\rho)^2}{1+\rho^2}\left[\frac{1}{2}\bar{q}^2 + \bar{q}e_L\right] \tag{6.44}$$

and

$$\beta_L^{oo*} = \frac{(1+\rho)}{1+\rho^2}\bar{q}\,, \tag{6.45}$$

where $\bar{q} = 1 - \frac{1-a}{a}(e_H - e_L)$. A contract that offers a fixed compensation of

$$\alpha_H^{oo*} = \frac{(1+\rho)^2}{1+\rho^2}\left[\bar{q}(e_H - e_L) - \frac{1}{2} - e_H\right] \tag{6.46}$$

and a bonus factor of

$$\beta_H^{oo*} = \frac{(1+\rho)}{1+\rho^2} \tag{6.47}$$

will only be accepted by H-agents. A comparison with equation (6.8) shows that the H-type is provided with efficient incentives in this menu of contracts. The employer is able to screen the agents and makes an expected profit of

$$E[\Pi^{oo*}(a)] = 2[ae_L + (1-a)e_H] + \frac{(1+\rho)^2}{2(1+\rho^2)}[a\bar{q}^2 + (1-a)] - c_o\,. \tag{6.48}$$

The H-agent can realize an expected utility of

$$E[U^H(\tilde{e}^{oo*}, \alpha_H^{oo*}, \beta_H^{oo*})] = \frac{(1+\rho)^2}{1+\rho^2}q(e_H - e_L) \tag{6.49}$$

by choosing the contract containing α_H^{oo*}, β_H^{oo*}, and one evaluation. The L-agent can only realize a utility of zero with the contract designed for him. Choosing the H-type's contract would leave him with a negative expected utility.

Two Evaluations

Once the principal decides to screen via a two-evaluation contract, his expected payoff is

$$\begin{aligned} E[\Pi^{tt}(a)] &= a[2(1-\beta_L^{tt})(\beta_L^{tt}+e_L)-\alpha_L^{tt}] \\ &\quad + (1-a)[2(1-\beta_H^{tt})(\beta_H^{tt}+e_H)-\alpha_H^{tt}] - c_t\,. \end{aligned} \tag{6.50}$$

The optimal compensation parameters for the L-agent's contract are

$$\alpha_L^{tt*} = -\bar{q}^2 - 2\bar{q}e_L \tag{6.51}$$

and

$$\beta_L^{tt*} = \bar{q}\,. \tag{6.52}$$

Only H-agents will choose a contract with the compensation parameters

$$\alpha_H^{tt*} = 2\bar{q}(e_H - e_L) - 1 - 2e_H \tag{6.53}$$

and

$$\beta_H^{tt*} = 1\,. \tag{6.54}$$

So, as in the previous section, the H-agent is provided with efficient incentives. His expected utility equals

$$E[U^H(\tilde{e}^{oo*}, \alpha_H^{oo*}, \beta_H^{oo*})] = \frac{(1+\rho)^2}{1+\rho^2} q(e_H - e_L)\,. \tag{6.55}$$

Screening the agents is possible and it generates an expected profit of

$$E[\Pi^{tt*}(a)] = 2[ae_L + (1-a)e_H] + a\bar{q}^2 - +(1-a) - c_t\,. \tag{6.56}$$

Separating Contracts with Unequal Evaluation Frequencies

As our particular interest is the choice of the evaluation frequency, we are concerned with the outcome of a contract where the principal screens the agents by offering contracts with different evaluation frequencies. Compensation parameters and the evaluation frequency are highly interlinked, so obviously, a screening by the evaluation frequency will—if this is possible—additionally imply an adjustment of the compensation variables to the number of evaluations. Specifically, the principal can either differentiate by evaluating the L-agent once and the H-agent twice, or he can decide to do the inverse.

One and Two Evaluations

When the L-type opts into a contract with one evaluation and the H-type chooses the contract regime in which he is evaluated twice, the employer's expected profit, given optimal effort choice, is characterized by

$$\begin{aligned} E[\Pi^{ot}(a)] = \;& a[(2-(1+\rho)\beta_L^{ot})e_L + (1+\rho)\beta_L^{ot} - (1+\rho^2)(\beta_L^{ot})^2 - \alpha_L^{ot} - c_o] \\ & +(1-a)[2(1-\beta_H^{ot})(\beta_H^{ot}+e_H) - \alpha_H^{ot} - c_t] \; . \end{aligned} \tag{6.57}$$

Maximizing (6.57) subject to both, the agents' participation and incentive compatibility constraints, yields the optimal compensation parameters. For the L-agent's contract they equal

$$\alpha_L^{ot*} = -\left(\frac{2(1+\rho)^2 q}{a(1-\rho)^2+(1+\rho)^2}\right)\cdot\left(\frac{(1+\rho^2)q}{a(1-\rho)^2+(1+\rho)^2}+e_L\right) \tag{6.58}$$

and

$$\beta_L^{ot*} = \frac{2(1+\rho)q}{a(1-\rho)^2+(1+\rho)^2} \; . \tag{6.59}$$

H-types will choose a contract with the parameters

$$\alpha_H^{ot*} = -\left(\frac{(1+\rho)^2 q}{a(1-\rho)^2+(1+\rho)^2}\right)\cdot\left(\frac{(1+\rho)^2 q}{a(1-\rho)^2+(1+\rho)^2}+2e_L\right) \tag{6.60}$$

and

$$\beta_H^{ot*} = \frac{(1+\rho)^2 q}{a(1-\rho)^2 + (1+\rho)^2}\,. \tag{6.61}$$

The employer can screen the agents by offering these two contracts and makes an expected profit of

$$\begin{aligned} E[\Pi^{ot*}(a)] = \; & 2[ae_L + (1-a)e_H] + \\ & \tfrac{(1+\rho)^2 q^2}{a(1-\rho)^2+(1+\rho)^2} - ac_o - (1-a)c_t\,. \end{aligned} \tag{6.62}$$

The H-agents receives a utility of

$$U^H(\tilde{e}^{ot*}, \alpha_H^{ot*}, \beta_H^{ot*}) = \frac{2(1+\rho)^2 q}{a(1-\rho)^2 + (1+\rho)^2}(e_H - e_L)\,. \tag{6.63}$$

Two and One Evaluations

To complete our analysis, we will now derive the optimal contract parameters for a contract where the L-agent chooses a contract with two evaluations and the H-agent a contract with only one evaluation. An employer who offers this type of contract, will make an expected profit of

$$\begin{aligned} E[\Pi^{to}(a)] \;&=\; a[2(1-\beta_L^{to})(\beta_L^{to} + e_L) - \alpha_L^{to} - c_t] \\ &\quad + (1-a)[(2-(1+\rho)\beta_H^{to})e_H + (1+\rho)\beta_H^{to}] \\ &\quad - (1-a)[(1+\rho^2)(\beta_H^{to})^2 + \alpha_H^{to} + c_o]\,. \end{aligned} \tag{6.64}$$

Now, the optimal compensation parameters for the L-agent's contract are

$$\alpha_L^{to*} = -\left(\frac{(1+\rho)^2 q}{2(1+\rho^2) - a(1-\rho)^2}\right)\cdot\left(\frac{(1+\rho)^2 q}{2(1+\rho^2) - a(1-\rho)^2} + 2e_L\right) \tag{6.65}$$

and

$$\beta_L^{to*} = \frac{(1+\rho)^2 q}{2(1+\rho^2) - a(1-\rho)^2}\,. \tag{6.66}$$

H-agents select a one-evaluation contract with the parameters

$$\alpha_H^{to*} = -\left(\frac{2(1+\rho)^2 q}{2(1+\rho^2) - a(1-\rho)^2}\right) \cdot \left(\frac{(1+\rho^2)q}{2(1+\rho^2) - a(1-\rho)^2} + e_L\right) \tag{6.67}$$

and

$$\beta_H^{to*} = \frac{2(1+\rho)q}{2(1+\rho^2) - a(1-\rho)^2}\,. \tag{6.68}$$

The employer can screen the agents by offering these two contracts and makes an expected profit of

$$\begin{aligned} E[\Pi^{to*}(a)] &= 2[ae_L + (1-a)e_H] + \\ &\quad \tfrac{(1+\rho)^2 q^2}{2(1+\rho^2) - a(1-\rho)^2} - ac_t - (1-a)c_o\,. \end{aligned} \tag{6.69}$$

Given this optimal contract, the H-agent's utility amounts to

$$E[U^H(\tilde{e}^{to*}, \alpha_H^{to*}, \beta_H^{to*})] = \frac{2(1+\rho)^2 q}{2(1+\rho^2) - a(1-\rho)^2}(e_H - e_L)\,. \tag{6.70}$$

Results

By analyzing the results of the preceding sections, we find that the two pooling contracts are always dominated by the corresponding contracts that screen via the compensation parameters only. These contracts give optimal incentives to the H-agent and leave him with a positive rent. Comparing profits of both types of contract for one and two evaluations yields

$$E[\Pi^{oP*}(a) - \Pi^{oo*}(a)] < 0 \Leftrightarrow -\frac{(1+\rho)^2}{2(1+\rho^2)}\frac{(1-a)^3}{a}(e_H - e_L)^2 < 0 \tag{6.71}$$

and

$$E[\Pi^{tP*}(a) - \Pi^{tt*}(a)] < 0 \Leftrightarrow -\frac{(1-a)^3}{a}(e_H - e_L)^2 < 0 \tag{6.72}$$

respectively.

Proposition 8 *A principal will never offer a pooling contract when he can differentiate agents using the compensation parameters and evaluation frequency.*

The proof directly follows from equations (6.71) and (6.72).

Whilst the exclusion of pooling contracts is a clear-cut result, a principal's decision between the different separating contracts depends on the parametrization of the model. For the comparison between the screening contract with one evaluation and the screening contract with two evaluations, we get

$$E[\Pi^{oo*}(a) - \Pi^{tt*}(a)] \gtreqless 0 \Leftrightarrow \frac{(1-\rho^2)}{2(1+\rho^2)}[(1-a) + a\bar{q}^2] \lesseqgtr c_t - c_o\,. \tag{6.73}$$

The respective profit differential when comparing this option with a screening contract that evaluates L-agents once and H-agents twice is

$$\begin{aligned} E[\Pi^{oo*}(a) - \Pi^{ot*}(a)] &\gtreqless 0 \Leftrightarrow \\ \tfrac{(1+\rho^2)}{2(1+\rho^2)}[(1-a) + a\bar{q}^2] - \tfrac{(1+\rho)^2 q^2}{a(1-\rho)^2+(1+\rho)^2} &\gtreqless (1-a)(c_o - c_t)\,. \end{aligned} \tag{6.74}$$

The comparison with the reverse choice of evaluation frequency leads us to

$$\begin{aligned} E[\Pi^{oo*}(a) - \Pi^{to*}(a)] &\gtreqless 0 \Leftrightarrow \\ \tfrac{(1+\rho^2)}{2(1+\rho^2)}[(1-a) + a\bar{q}^2] - \tfrac{(1+\rho)^2 q^2}{2(1+\rho^2)-a(1-\rho)^2} &\gtreqless a(c_o - c_t)\,. \end{aligned} \tag{6.75}$$

The profit differential for a contract that screens, whilst all agents are evaluated twice, and a contract that evaluates L-agents once and H-agents twice, leads us to the following inequality

$$\begin{aligned} E[\Pi^{tt*}(a) - \Pi^{ot*}(a)] &\gtreqless 0 \Leftrightarrow \\ a\bar{q}^2 + (1-a) - \tfrac{(1+\rho)^2 q^2}{a(1-\rho)^2+(1+\rho)^2}) &\gtreqless a(c_t - c_o)\,. \end{aligned} \tag{6.76}$$

Furthermore, the behavior of the differential for this contract and a contract that evaluates L-agents twice and H-agents once can be described as

$$\begin{aligned} E[\Pi^{tt*}(a) - \Pi^{to*}(a)] &\gtreqless 0 \Leftrightarrow \\ a\bar{q}^2 + (1-a) - \tfrac{(1+\rho)^2 q^2}{2(1+\rho^2) - a(1-\rho)^2}) &\gtreqless (1-a)(c_t - c_o)\,. \end{aligned} \tag{6.77}$$

Finally, the trade-off between the two contracts which differentiate via evaluation frequency is characterized by

$$\begin{aligned} E[\Pi^{to*}(a) - \Pi^{ot*}(a)] &\gtreqless 0 \Leftrightarrow \\ \tfrac{(1+\rho)^2(1-\rho)^2 q^2 (1-2a)}{(2(1+\rho^2) - a(1-\rho)^2)(a(1-\rho)^2 + (1+\rho)^2)} &\gtreqless (c_t - c_o)(1-2a)\,. \end{aligned} \tag{6.78}$$

It is easy to see that profits are equal, when the shares of L- and H-agents are equal, i.e. $a = 0.5$.

As the inequalities (6.71) to (6.78) crucially depend on the parametrization of the model, figures 6.1 to 6.5 visualize the profits generated by different contract choices for several values of e_H, e_L, c_o, and c_t, or, to be more precise, for values of the relevant parameter differences $e_H - e_L$ and $c_t - c_o$.[5] The figures show variations of profit with respect to the share of L-agents, which is a, and ρ, the memory parameter of the principal.

For all parameterizations, depending on the values of a and ρ, either one-evaluation or two-evaluation screening contracts are being offered by the principal. Contracts screening via evaluation frequency are always dominated by one of the former screening contracts using only one evaluation frequency for both agents. This is intuitive as the principal can successfully screen agents via compensation parameters while choosing the optimal evaluation frequency independently of his screening activity. Anyhow, as we can see in figures 6.2 and 6.3 the frequency that is chosen in the screening contract still depends on the distribution of types in

[5] The different parameterizations are chosen in a "sensitive" way, i.e. we have chosen parameterizations that illustrate particularly well the trade-offs that can occur in the model. Also, the choice of $e_H - e_L$ has been made such, that negative bonus payments can be excluded with certainty.

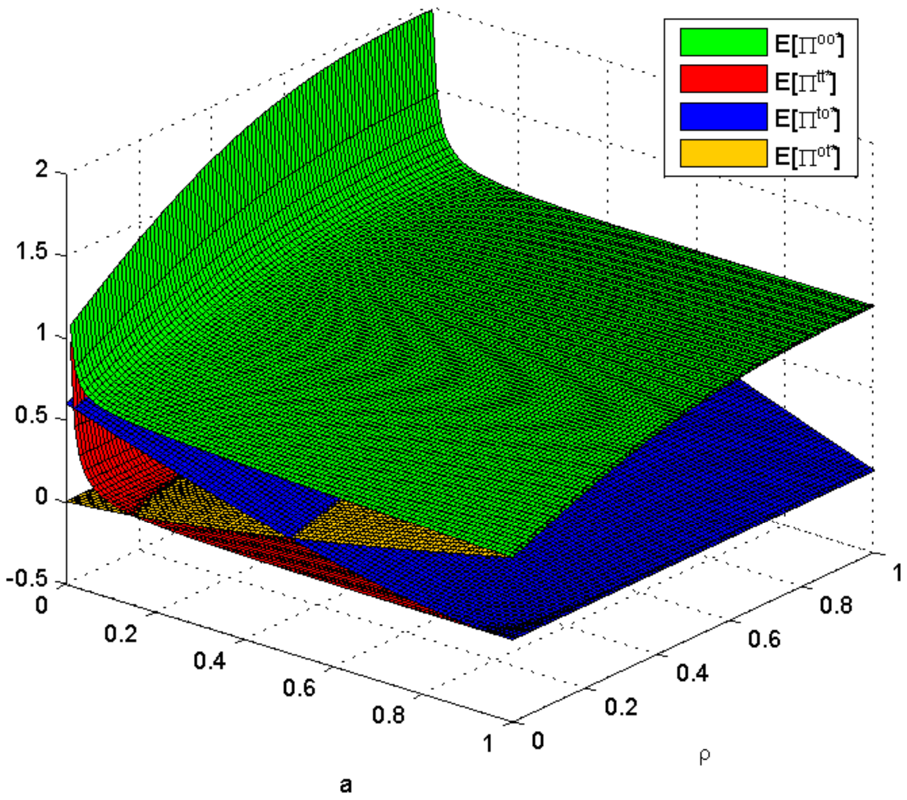

Figure 6.1: Screening Profits (Parametrization: $e_H - e_L = 0.1, c_t - c_o = 1$)

the population. When the share of L-agents a is rather low, two evaluations pay off more than when only few H-agents are employed. Furthermore, the graphics suggest that if the memory parameter ρ impacts the principal's contract choice it does so in only one way: Higher values of ρ make one-evaluation contracts more attractive, which is a reasonable result.

For, so to say, "extreme" parameterizations, the choice of evaluation frequency becomes independent of a and ρ (figures 6.1 and 6.5). For instance, when the heterogeneity is low but additional evaluations are costly (Figure 6.1), only one-evaluation screening contracts will be offered. We see that with lower values of $c_t - c_o$, contracts with more evaluations become more profitable. When $e_H - e_L$ and $c_t - c_o$ are equal (Figures 6.2 and 6.3), lower parameter values of the former

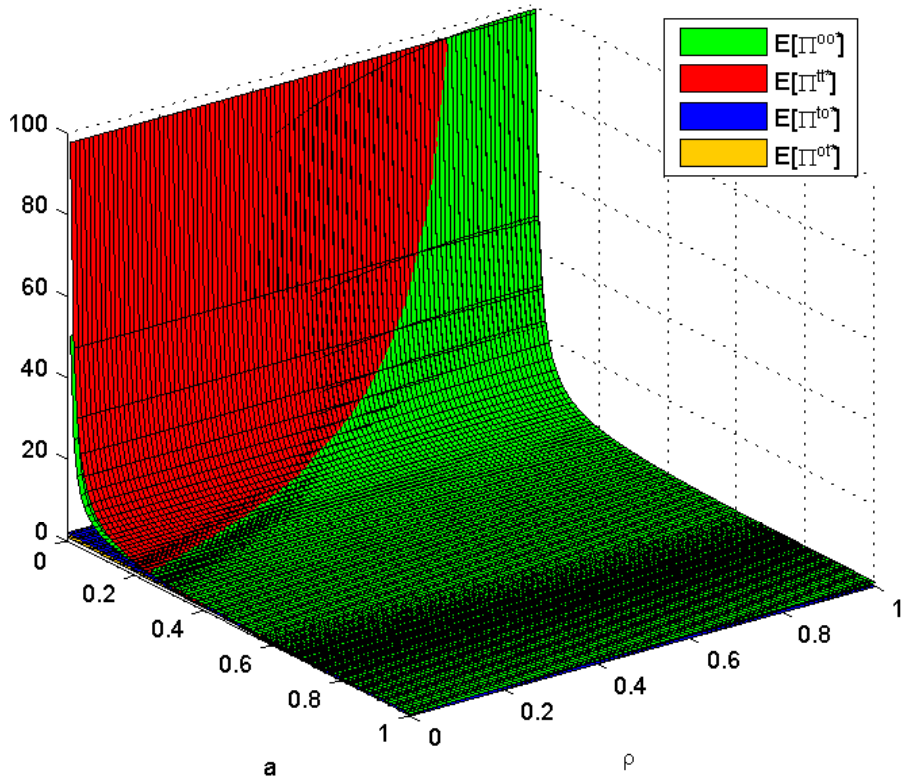

Figure 6.2: Screening Profits (Parametrization: $e_H - e_L = 1, c_t - c_o = 1$)

make two-evaluation screening contracts relatively more attractive for low values of ρ. Reducing the additional cost of an evaluation to a value close to zero makes the two-evaluation screening contracts the only ones being offered for all values of a and ρ (Figure 6.5). Altogether, we can state that there are only two remaining contract options to be considered by the principal. He will always screen the heterogenous agents using solely the compensation parameters.

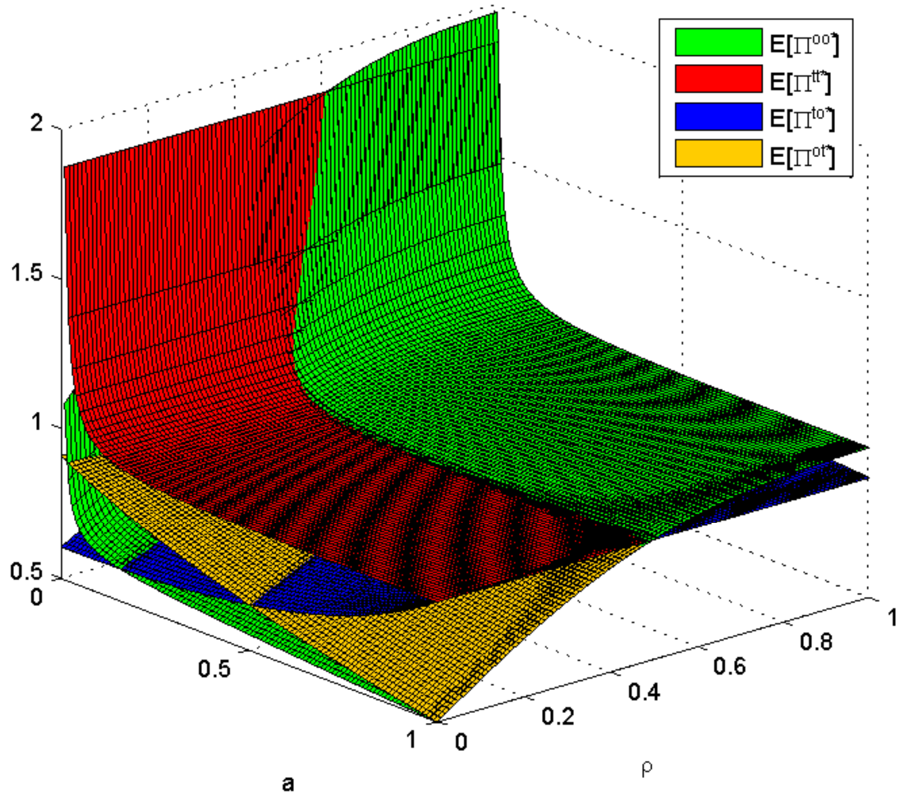

Figure 6.3: Screening Profits (Parametrization: $e_H - e_L = 0.1, c_t - c_o = 0.1$)

6.4.2 Restriction to Equal Bonus Parameters

Unequal Evaluation Frequencies

In the previous discussion of results, it has turned out that screening contracts differentiating with respect to all compensation variables are dominated by contracts that screen by the use of the compensation parameters only. The reason for this is that the principal can already screen the agents by adjusting compensation and he will therefore choose only one optimal evaluation frequency which is identical for both types of agents. So screening with respect to evaluation frequency is not optimal for the principal in this setup. But what would happen if ratings and bonus payments were public? For instance, it is plausible to think that for

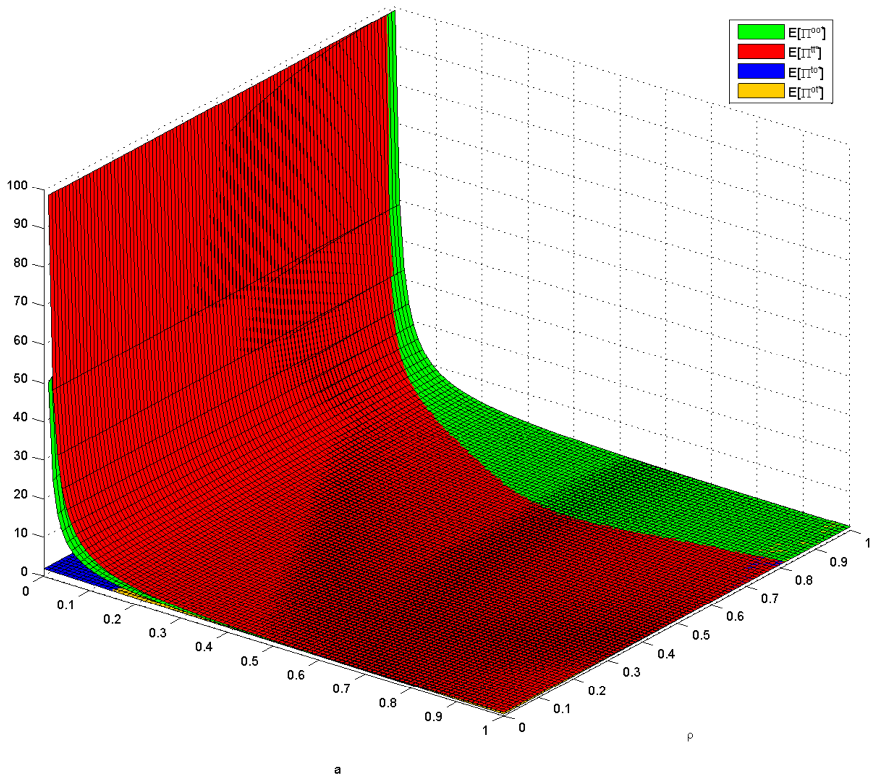

Figure 6.4: Screening Profits (Parametrization: $e_H - e_L = 1, c_t - c_o = 0.01$)

reasons of equity, agents cannot be paid a lower bonus than a peer for having rendered identical performance. In the following two sections, we will therefore analyze work contracts where a differentiation can be made according to the fixed wage as well as according to the evaluation frequency but for equal performance, bonuses must be the same for all agents.

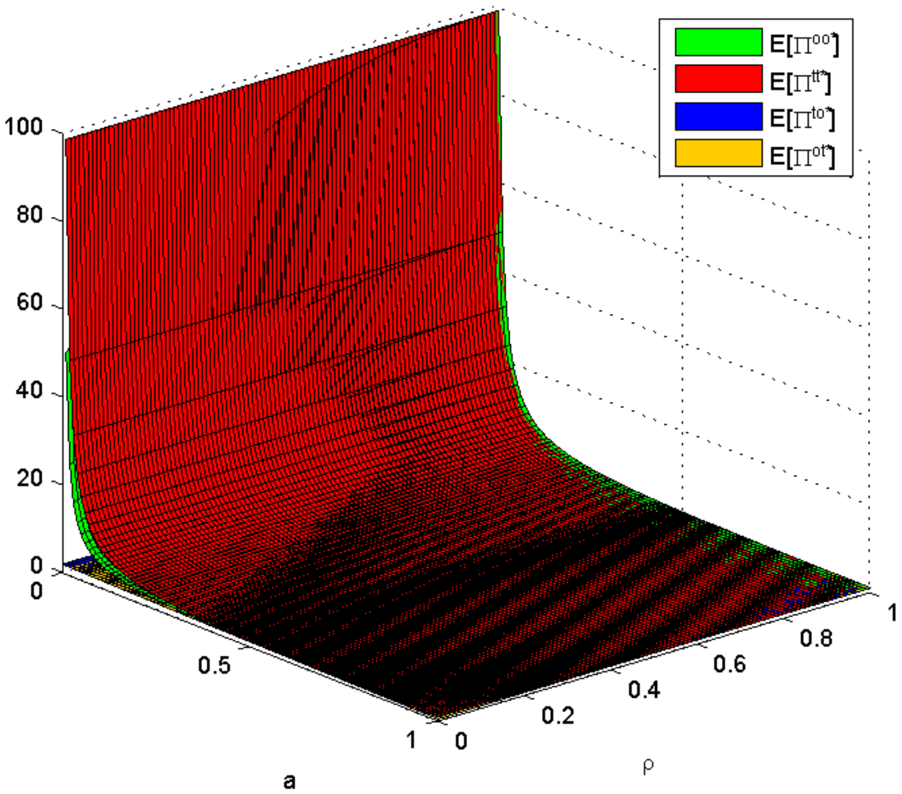

Figure 6.5: Screening Profits (Parametrization: $e_H - e_L = 1, c_t - c_o = 0.0001$)

One and Two Evaluations

The principal's expected profit from a screening contract with one evaluation for the L-type and two evaluations for the H-type now is characterized by

$$\begin{aligned} E[\hat{\Pi}^{ot}(a)] = \quad & a[(2-(1+\rho)\hat{\beta}^{ot})e_L + (1+\rho)\hat{\beta}^{ot} - (1+\rho^2)(\hat{\beta}^{ot})^2 - \hat{\alpha}_L^{ot} - c_o] \\ & +(1-a)[2(1-\hat{\beta}^{ot})(\hat{\beta}^{ot} + e_H) - \hat{\alpha}_H^{ot} - c_t]\,. \end{aligned} \tag{6.79}$$

Note that the bonus parameter $\hat{\beta}^{ot}$ now is set equal for both types of agents. Maximizing (6.79) subject to both, the agents' participation and incentive compatibility constraints, yields the optimal compensation parameters. The optimal bonus parameter for both contracts is

$$\hat{\beta}^{ot*} = \frac{2 - a(1-\rho) - (1+\rho)(1-a)(e_H - e_L)}{2 - a(1-\rho^2)} \,. \tag{6.80}$$

L-agents will be willing to choose the one-evaluation contract if the fixed wage offered equals

$$\hat{\alpha}_L^{ot*} = -1/2(1+p^2)(\hat{\beta}^{ot*})^2 - (1+p)\hat{\beta}^{ot*}e_L \,. \tag{6.81}$$

The optimal fixed wage component in the two-evaluation contract designated to the H-types is

$$\hat{\alpha}_H^{ot*} = -(\hat{\beta}^{ot*})^2 - 2e_H\hat{\beta}^{ot*} + (1+p)\hat{\beta}^{ot*}(e_H - e_L) \,. \tag{6.82}$$

The employer's expected profit from this menu of contracts amounts to

$$\begin{aligned} E[\hat{\Pi}^{ot*}(a)] = \;& 2[ae_H + (1-a)e_L] + \\ & \tfrac{[2-a(1-\rho)-(1+\rho)(1-a)(e_H-e_L)]^2}{2[2-a(1-p^2)]} - ac_o - (1-a)c_t \,. \end{aligned} \tag{6.83}$$

Whereas the H-agents receive a utility of

$$U^H(\tilde{e}^{ot*}, \hat{\alpha}_H^{ot*}, \hat{\beta}^{ot*}) = \hat{\beta}^{ot*}(1+\rho)(e_H - e_L) \,, \tag{6.84}$$

the utility of the L-agents is zero.

Two and One Evaluations

Analogously to the previous section, we derive the optimal contract parameters for a contract where the L-agent is evaluated twice and the H-agent only once. The employer's expected profit will be

$$\begin{aligned} & E[\hat{\Pi}^{to}(a)] = a[2(1-\hat{\beta}^{to})(\hat{\beta}^{to} + e_L) - \hat{\alpha}_L^{to} - c_t] \\ & +(1-a)[(2-(1+\rho)\hat{\beta}^{to})e_H + (1+\rho)\hat{\beta}^{to} - (1+\rho^2)(\hat{\beta}^{to})^2 - \hat{\alpha}_H^{to} - c_o] \,. \end{aligned} \tag{6.85}$$

The optimal bonus parameter solving the principal's maximization problem is

$$\hat{\beta}^{to*} = \frac{a(1-\rho) + (1+\rho) - 2(1-a)(e_H - e_L)}{a(1-\rho^2) + (1+\rho^2)} \,. \tag{6.86}$$

L-agents are willing to be evaluated twice if the fixed compensation component equals

$$\hat{\alpha}_L^{to*} = -(\hat{\beta}^{to*})^2 - 2e_L\hat{\beta}^{to*} \tag{6.87}$$

and H-agents prefer to be evaluated only once receiving a fixed wage of

$$\hat{\alpha}_H^{to*} = -1/2(1+\rho^2)(\hat{\beta}^{to*})^2 - (1+\rho)\hat{\beta}^{to*}e_H + 2\hat{\beta}^{to*}(e_H - e_L)\,. \tag{6.88}$$

The employer can screen the agents by offering these two contracts and makes an expected profit of

$$\begin{aligned} E[\hat{\Pi}^{to*}(a)] = \; & 2[ae_H + (1-a)e_L] + \\ & \tfrac{[a(1-\rho)+(1+\rho)-2(1-a)(e_H-e_L)]^2}{2[a(1-\rho^2)+(1+\rho^2)]} - ac_t - (1-a)c_o \;. \end{aligned} \tag{6.89}$$

Given this optimal contract, the H-agent's utility amounts to

$$E[U^H(\tilde{e}^{to*}, \hat{\alpha}^{to*}, \hat{\beta}_H^{to*})] = 2\hat{\beta}^{to*}(e_H - e_L) \;. \tag{6.90}$$

Results

To identify the most profitable contract in a situation with equal bonus parameters for both types of agents we must compare several outcomes: Those of the two screening contracts derived in the previous section (equations (6.83) and (6.89)) and those of the two pooling contracts (equations (6.35) and (6.35)). In contrast to the results where a differentiation with respect to bonus parameters is possible, we find that screening via the choice of evaluation frequency can now be optimal.

Comparing profits from the two pooling contracts yields

$$E[\Pi^{oP*}(a) - \Pi^{tP*}(a)] \gtreqless 0 \Leftrightarrow -\frac{(1-\rho)^2}{2(1+\rho^2)} q^2 \gtreqless c_o - c_t \,. \tag{6.91}$$

The profit from the pooling contract with one evaluation compared to those from the two screening contracts yields the inequalities

$$\begin{array}{rcc} E[\Pi^{oP*}(a) - \hat{\Pi}^{ot*}(a)] & \gtreqless & 0 \quad \Leftrightarrow \\ \frac{(1+\rho)^2}{2(1+\rho^2)} q^2 - \frac{[2-a(1-\rho)-(1+\rho)(1-a)(e_H-e_L)]^2}{2[2-a(1-\rho^2)]} & \gtreqless & (1-a)(c_o - c_t) \end{array} \tag{6.92}$$

and

$$\begin{array}{rcc} E[\Pi^{oP*}(a) - \hat{\Pi}^{to*}(a)] & \gtreqless & 0 \quad \Leftrightarrow \\ \frac{(1+\rho)^2}{2(1+\rho^2)} q^2 - \frac{[a(1-\rho)+(1+\rho)-2(1-a)(e_H-e_L)]^2}{2[a(1-\rho^2)+(1+\rho^2)]} & \gtreqless & a(c_o - c_t)\,. \end{array} \tag{6.93}$$

Pooling with two evaluations dominates the screening contracts when the left-hand side of the following two inequalities outbalances the respective right-hand side:

$$\begin{array}{rcc} E[\Pi^{tP*}(a) - \hat{\Pi}^{ot*}(a)] & \gtreqless & 0 \quad \Leftrightarrow \\ q^2 - \frac{[2-a(1-\rho)-(1+\rho)(1-a)(e_H-e_L)]^2}{2[2-a(1-\rho^2)]} & \gtreqless & a(c_t - c_o)\,, \end{array} \tag{6.94}$$

$$\begin{array}{rcc} E[\Pi^{tP*}(a) - \hat{\Pi}^{to*}(a)] & \gtreqless & 0 \quad \Leftrightarrow \\ q^2 - \frac{[a(1-\rho)+(1+\rho)-2(1-a)(e_H-e_L)]^2}{2[a(1-\rho^2)+(1+\rho^2)]} & \gtreqless & (1-a)(c_t - c_o)\,. \end{array} \tag{6.95}$$

Again, we recur to a parametrization to illustrate changes of the model outcome with respect to variations of the parameter set. To allow a comparison with the results in a situation where a differentiation of all choice parameters is possible, the following figures are based on the same parametrization as Figures 6.1 to 6.5.

In contrast to the parameterizations in Section 6.4.1, Figures 6.6 to 6.10 indicate that, in the restricted model version, all four possible contract regimes can be optimal. The choice of the principal varies in all parameters. Especially, variations in a and ρ are multifaceted in this modified setup. Figure 6.9 illus-

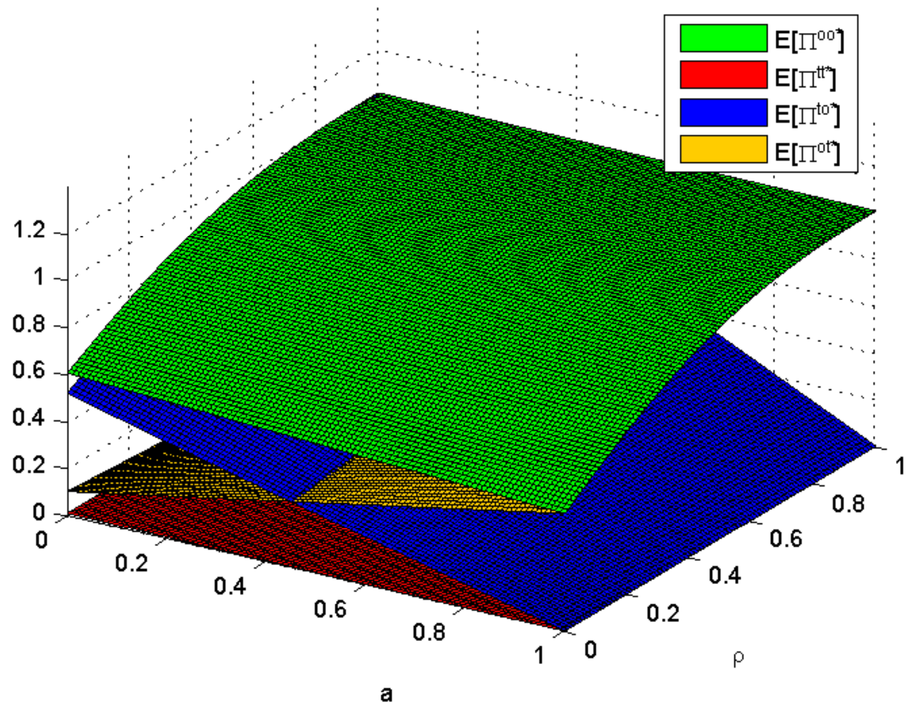

Figure 6.6: Pooling and Screening Profits (Parametrization: $e_H - e_L = 0.1$, $c_t - c_o = 1$)

trates this particularly well. Depending on a and ρ, four different options can be optimal with this specific parametrization. Screening the agents with respect to evaluation frequency is favorable when there are relatively few L-agents, i.e. when a is low (Figures 6.7 to 6.10). Otherwise, pooling contracts will be chosen. Increases in the memory parameter ρ lower the benefit from more evaluations and, therefore, they can lower the number of evaluations whenever they induce a change from a screening to a pooling contract. A switch of the screening contracts is also possible due to a change in ρ (see Figure 6.7, 6.9, and 6.10).

Variations in evaluation cost clearly affect the outcome. For a very high cost of an additional evaluation, the outcome is identical to the one in the unrestricted model. Figure 6.6 indicates that only contracts with one evaluation will be chosen. The difference is that, in the restricted model this is a pooling con-

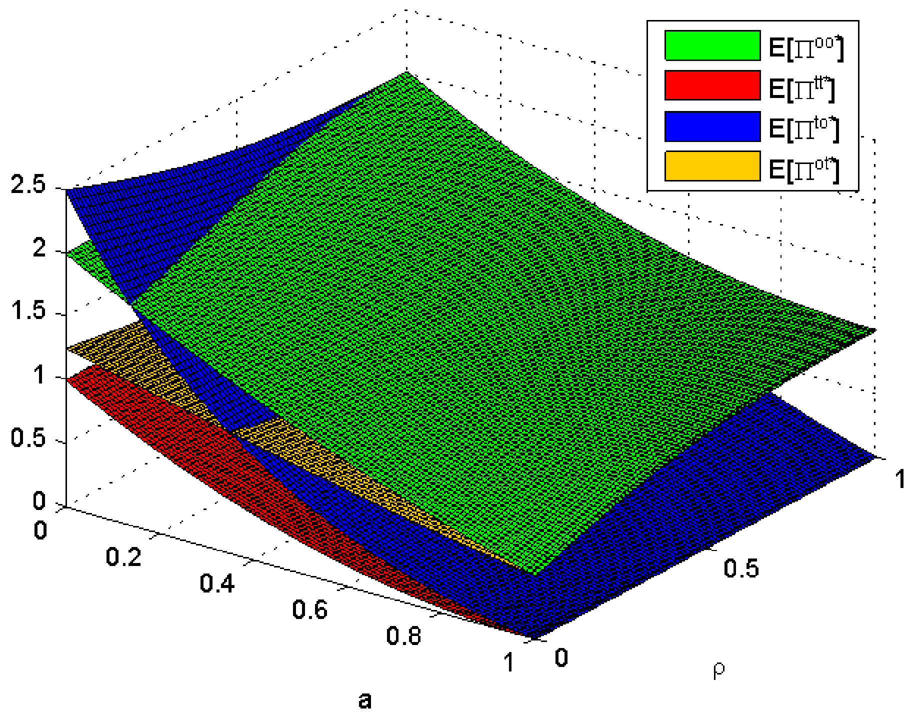

Figure 6.7: Pooling and Screening Profits (Parametrization: $e_H - e_L = 1$, $c_t - c_o = 1$)

tract, whereas, in Section 6.4.1, a screening contract with one evaluation for both agents was feasible and optimal. When additional evaluations are relatively cheap compared to the degree of heterogeneity $e_H - e_L$ (Figure 6.10), not all agents will be evaluated twice. This stands in contrast to Section 6.4.1 as the principal now uses the differentiation with respect to evaluation frequency to screen the agents. We can resume that, given the restriction of the choice of the bonus factor, all four remaining contract options must be considered by the principal. Reducing the number of instruments for differentiation aggravates the principal's decision problem. He faces a trade-off between optimal screening on the one hand and optimal evaluation frequency on the other hand.

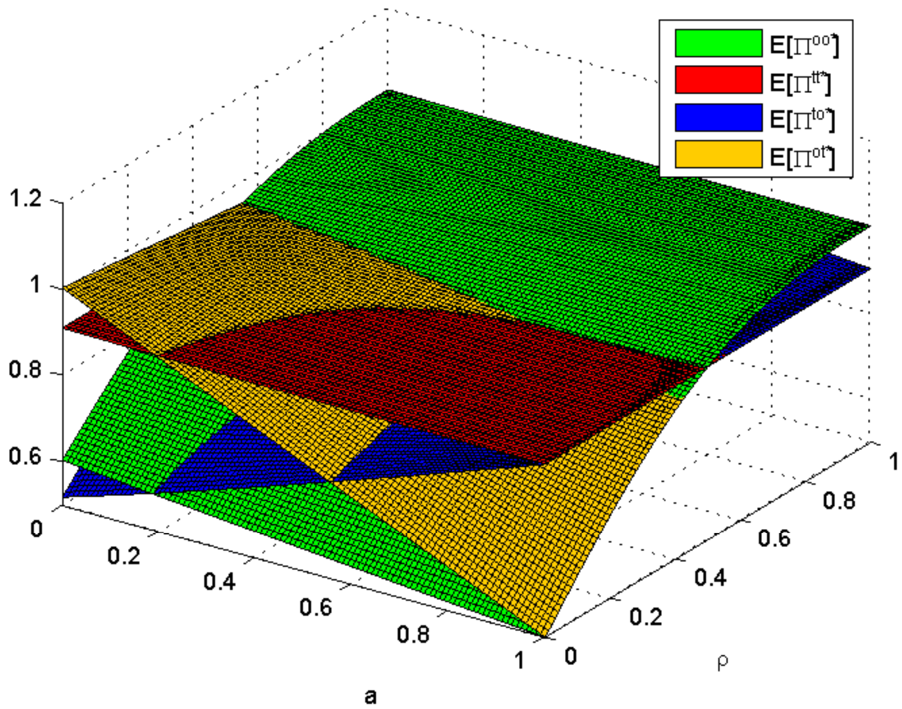

Figure 6.8: Pooling and Screening Profits (Parametrization: $e_H - e_L = 0.1$, $c_t - c_o = 0.1$)

6.5 Conclusion

We have proposed a model of performance-based contracts in the presence of asymmetric information concerning agents' productivity and under the assumption of bounded memory on behalf of the principal who conducts evaluations. Within different sets of feasible contracts, we have formalized optimal pooling and separating contracts.

Under symmetric information, the choice of the optimal evaluation frequency is independent of the agent's type. Though agents are risk-neutral, we do not confirm the result from Chapter 3 that it is never optimal for the principal to evaluate more often than at the end of the contracting period. The difference in the results is caused by the quadratic effort cost, adapted from Moen and Rosén

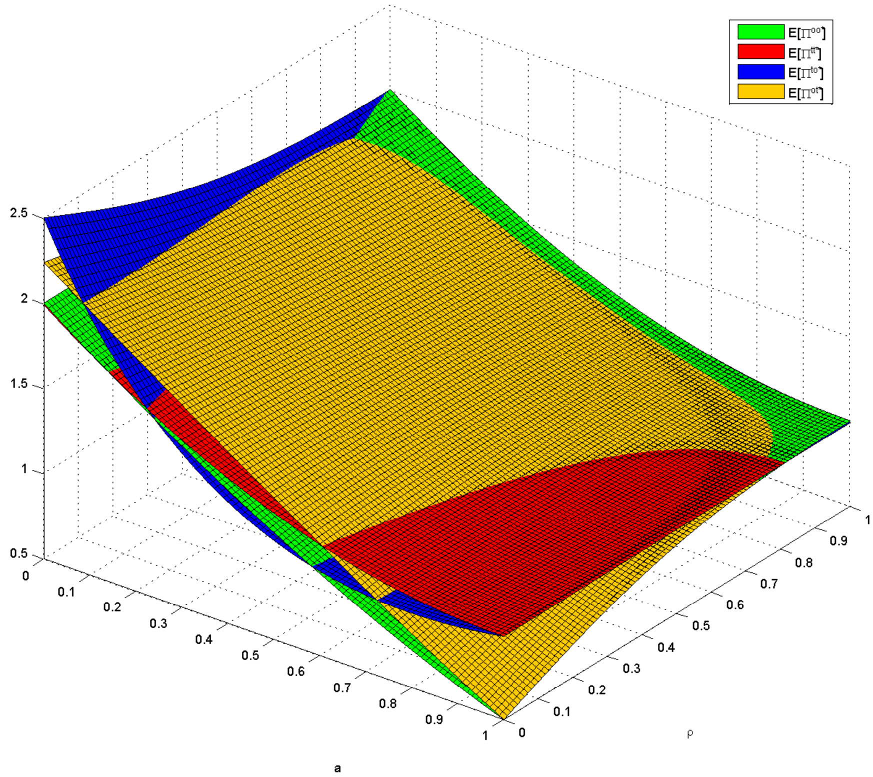

Figure 6.9: Pooling and Screening Profits (Parametrization: $e_H - e_L = 1$, $c_t - c_o = 0.01$)

(2005). Whereas, in Chapters 3 and 4, investing in first-period effort is always as costly to the agent as investing in second-period effort, no matter the current level of investment, this is not true in a setup with quadratic effort cost.

Furthermore, we find that screening by the use of different evaluation frequencies is only worthwhile when other screening possibilities are not available. If agents can be screened by the help of compensation parameters, the optimal choice of evaluation frequency is the same as under symmetric information. Hence, both agents will be evaluated at the same frequency due to the independence of the optimal evaluation frequency from the type of agent. As opposed to a setup with homogenous agents, the composition of the population of workers

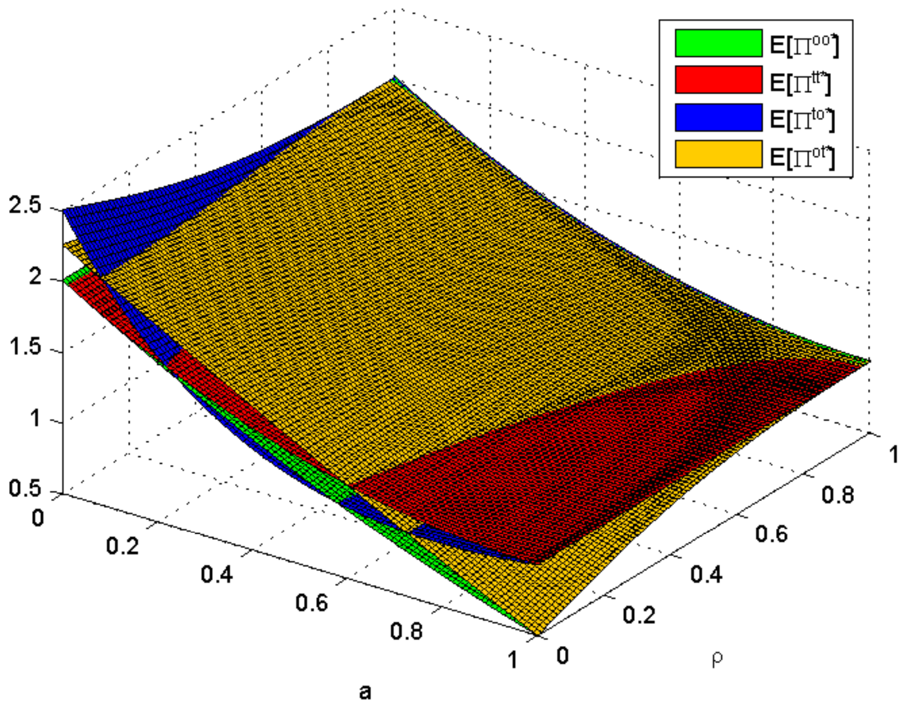

Figure 6.10: Pooling and Screening Profits (Parametrization: $e_H - e_L = 1$, $c_t - c_o = 0.0001$)

has an impact on optimal evaluation frequency. The more L-types are present in the work force, the less evaluations should be conducted. The optimal evaluation frequency will only be chosen differently for the two types if there are no other screening tools at hand which can be used at an acceptable cost. If paying agents different bonuses for the same performance causes feelings of inequity within a work force and if this again is costly due to a negative effect on motivation, a differentiation with respect to bonus payments actually might be impossible.

This chapter gives further theoretical foundations for the choice of the frequency of employee evaluations. It can explain why evaluation frequency might differ between units of a firm or between organizations but maybe not between employees performing similar tasks. Anecdotic evidence already suggests, that for instance consultants—who are a good example for highly qualified employees—are being evaluated more often compared to other business professionals. However, we have not yet observed a variation of evaluation frequency within one group of employees doing a similar job, which is rather in line with our results. An empirical study of evaluation frequency would be helpful for a further understanding of this issue. It could help to investigate whether indirect indicators for productivity such as skill level and the position within an organization can successfully explain some of the variation in the practiced evaluation frequency.

6.6 Appendix

6.6.1 Derivation of the Aggregate Cost Function

To derive the cost function in equation (6.11) we first expand the expected value of (6.2) by $-(1+\rho)e^0$:

$$\begin{aligned} \tilde{e} - (1+\rho)e^0 &= \rho e_1 + e_2 - (1+\rho)e^0 \\ \Leftrightarrow \tilde{e} - (1+\rho)e^0 &= \rho(e_1 - e^0) + e_2 - e^0 \,. \end{aligned} \tag{6.96}$$

Equation (6.6) allows us to substitute $e_2 - e^0$ by $\frac{1}{\rho}(e_1 - e^0)$. Hence, we can simplify the right-hand side of (6.96) as follows:

$$\begin{aligned} \frac{1+\rho^2}{\rho}(e_1 - e^0) &= \tilde{e} - (1+\rho)e^0 \\ \Leftrightarrow e_1 - e^0 &= \frac{\rho}{1+\rho^2}(\tilde{e} - (1+\rho)e^0) \,. \end{aligned} \tag{6.97}$$

Substituting (6.6) into (6.97) gives us

$$e_1 - e^0 = \frac{1}{1+\rho^2}(\tilde{e} - (1+\rho)e^0) \,. \tag{6.98}$$

Substituting (6.97) and (6.98) into the cost function from (6.3) yields

$$\begin{aligned} C(\tilde{e}) &= \frac{1}{2}(e_1 - e^0)^2 + \frac{1}{2}(e_2 - e^0)^2 \\ &= \frac{1}{2}\left[\frac{\rho}{1+\rho^2}(\tilde{e} - (1+\rho)e^0)\right]^2 + \frac{1}{2}\left[\frac{1}{1+\rho^2}(\tilde{e} - (1+\rho)e^0)\right]^2 \\ &= \frac{(\tilde{e} - (1+\rho)e^0)^2}{2(1+\rho^2)}. \end{aligned} \tag{6.99}$$

Chapter 7

Conclusion

Summary

This work has discussed the role of the frequency with which performance evaluations take place. In a situation where the quality of information is decreasing over time—we explained this with the imperfectness of human memory—we analyzed the impact of several influential factors such as risk attitudes, wealth, or heterogeneity with respect to productivity on the conduct of evaluations.

In Chapter 3, we approached the issue in a setup where agents are risk-neutral or wealth-constrained. We found that the principal will not make use of more frequent evaluations to counterbalance his forgetfulness when the agent's wealth is unbounded. The consideration of wealth-constrained agents weakens this outcome and we established that the optimal evaluation frequency increases with the degree of forgetfulness, with falling fixed cost of the evaluation, and with the tightness of the agent's wealth constraint. As a main result of Chapter 4, the introduction of risk aversion on the part of the agent leads to a trade-off concerning the principal's choice of evaluation frequency. Though they are costly, in a model with risk aversion, more evaluations can even enhance efficiency through the elimination of risk within the evaluation process. The extension to a duopoly setup presented in Chapter 5 does not change the decision on evaluation frequency. Possibly, a competitive advantage for the principal who has the better memory capacity emerges. His share of the surplus increases the more the competitor forgets.

In the presence of heterogeneity concerning agents' productivity (Chapter 6) the optimal evaluation frequency potentially varies with the distribution of types in the population of agents. Under symmetric information, the choice regarding the optimal evaluation frequency is independent of the agent's type. In this chapter, we also introduced a quadratic cost function for the effort exertion of agents and found that this creates a trade-off concerning the evaluation frequency, even though agents are assumed to be risk-neutral. This, of course, shows that the result obtained in Chapter 3, i.e. that increasing evaluation frequency is wasteful when the agents are risk-neutral, only holds under very particular assumptions. With asymmetric information, we found that screening by the use of different evaluation frequencies is worthwhile if, and only if, other screening possibilities are not available. This suggests that sorting employees by different evaluation frequencies should only be observed if restrictions concerning the design of labor contracts are at work. For example, the transparency of a bonus system can prevent the differentiation of bonus payments within a work force because of equity considerations.

Our first results suggested that evaluations are a necessary evil and should be avoided as long as it is possible to do so. The setup generating this result was a puristic one: no risk aversion, linear effort cost, and agents who are homogenous in productivity and who face no wealth constraint. The following modifications all come along with deviations from and restrictions to this simple framework and the results they produce all have one thing in common. With rising complexity of the situation that is modeled, the decision on the optimal contract design, specifically the one on the optimal evaluation frequency, also becomes more complex. Common sense tells us that individuals *are* risk-averse, we know that they *do* face wealth constraints and that labor forces are heterogenous with respect to productivity. As already addressed in the introduction, a model of frequency decisions cannot yield a magic formula allowing to calculate the optimal evaluation frequency for specific situations. However, what the models we presented *can* do is to show that the choice of evaluation frequency within an organization should be made thoroughly, taking into account the specific environment.

Alternative Interpretations

Throughout the text, we have argued that the quality of information decreases over time and this is caused by the bounded capacity of the supervisor's memory. But our model can be applied to any situation where investment into measurement technology, a technology straightening measurement distortions, leads to more precise performance information.

The loss of information could as well be induced by the turnover of the supervisors. If an employee's superior quits the firm, the information he has gathered on his subordinate's performance will be lost. So more frequent evaluations, documenting the judgment, assure that this information stays within the firm and will be used for the determination of the bonus. Another possibility is interpreting the memory parameter as a discount factor, depreciating the value of the output as it has been done in Section 3.5. In the following section, we will also discuss evaluations in the light of hyperbolic discounting.

Different Functions of Evaluations

Coming back to the example of evaluations in the course of a Ph.D. program, two more important aspects concerning evaluation frequency should be considered. First of all, it is important to note that evaluations have an impact on two dimensions: past and future. In our setup, we only discussed the backwardoriented component of evaluations, but in addition, the frequency of their conduct plays a future-oriented role. If the feedback given in an evaluation is a means for players to learn, more evaluations should accelerate such a learning process. Not only individuals can learn about their fit for the current job or their overall ability (as modeled in Lazear; 1990), feedback also has a value for the firm, as individuals can optimize the allocation of effort to tasks more quickly, thanks to the information gained from the feedback process. This mechanism becomes particularly evident for the example of Ph.D. students. If their supervisor evaluates them at a frequent rate, this will allow students to identify promising aspects for research at an early stage, preventing them from putting too much effort into directions that are actually on the wrong track.

A further aspect of the rate at which evaluations take place is a motivational one that goes beyond the bonus payment as such. When individuals know that an evaluation is approaching, this will usually motivate them, boosting performance just before the review. This all too human behavior is called procrastination. In the individual perception, working hard from tomorrow on is preferable to starting to work hard today. Again, the Ph.D. student is our real life example par excellence. If the only evaluation he expects to receive is the assessment of his thesis in one, two, or more years, it might be quite difficult to continually work on his research from the start. A doctoral program that schedules regular workshops, during which the first drafts of the thesis are discussed and evaluated by the supervisor, will give a stronger motivation to produce results already at an early stage.

Who Wants Frequent Evaluations?

In general, employees as well as students dislike evaluations, whereas it is common belief that supervisors and teachers are fond of them. But the present work suggests that this view is too simplistic. There are many reasons why those who *are* evaluated should wish to be evaluated even more often. They receive important information about their actual performance which can help them to do an even better job in the future. Furthermore, doing the evaluation close to the time span in which the work actually has been done will lead to a more precise judgment, and disagreements can probably be settled more easily when the evaluation is close to the situation in question. When the labor force of an organization is relatively mobile, more evaluations maintain the quality of performance information, as the information loss due to turnover is lower. Last but not least, individuals who are procrastinators—and who are unhappy about this—should appreciate an increased number of evaluations.

For employers, as well as teachers and professors, evaluations first of all mean one thing: work. The benefits of evaluations mainly are indirect: an increased motivation of the supervised to perform, a reduction of the risk imposed on them, and an enhanced learning process. So, frequent evaluations should rather be a sign of an employer's attractiveness than a blemish. The results from Chapter 6 and also observations of business practice, indicate that employers do not want to increase evaluation frequency for the use of screening.

Future Research Issues

All in all, there have been very few attempts to explain variations of the frequency of certain actions in economic behavior. This is particularly true for the study of the choice of evaluation frequency of employees. This work is only another step towards a proper understanding of the role of evaluations and their potential within firms and other organizations. It would be of further interest to study the consequences of combining the different models presented in Chapters 3 to 6 with the approach proposed by Lazear (1990). This could be particularly fruitful, as the literature on evaluations is, in a certain sense, the link between the literature on performance pay and the research on feedback. Furthermore, an empirical verification, particularly of whether evaluations are used for information transmission in the sense of a feedback or rather for the improvement of the quality of performance information, would be of great interest. This would help to indicate in how far evaluations are used to enhance learning and how much they account for incentives provided in a firm.

Bibliography

Akemann, M. and Kanczuk, F. (2003). Term limits and term lengths, *Working Paper*.

Akerlof, G. (1970). The market for 'lemons': Quality uncertainty and the market mechanism, *Quarterly Journal of Economics* **84**: 488–500.

Alchian, A. A. and Demsetz, H. (1972). Production, information costs, and economic organization, *American Economic Review* **62**(5): 777–795.

Azariadis, C. (1975). Implicit contracts and underemployment equilibria, *Journal of Political Economy* **83**(6): 1183–1202.

Backes-Gellner, U., Lazear, E. P. and Wolff, B. (2001). *Personalökonomik*, Schäffer Poeschel.

Baker, G. P. (1992). Incentive contracts and performance measurement, *Journal of Political Economy* **100**(3): 598–614.

Baker, G. P. (2000). The use of performance measures in incentive contracting, *American Economic Review* **90**(2): 415–420.

Baker, G. P., Gibbons, R. and Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts, *Quarterly Journal of Economics* **109**(4): 1125–56.

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

Budde, J. (2006). Distorted performance measurement and relational contracts, *SFB/TR 15, GESY, Discussion Paper* (203).

Bull, C. (1987). The existence of self-enforcing implicit contracts, *Quarterly Journal of Economics* **102**(1): 147–159.

Coase, R. H. (1937). The nature of the firm, *Economica* **4**(16): 386–404.

D'Andrade, R. G. (1974). Memory and the assessment of behavior, *in* T. Blalock (ed.), *Measurement in the social sciences*, Aldine-Atherton, chapter 5, pp. 139–186.

Frey, B. S. (2005). "Just forget it": Memory distorsion as bounded rationality, *Mind & Society* **4**(1): 13–25.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames, *Review of Economic Studies* **38**(113): 1–12.

Gibbs, M., Merchant, K. A., van der Stede, W. A. and Vargus, M. E. (2004). Determinants and effects of subjectivity in incentives, *The Accounting Review* **79**(2): 409–436.

Grossman, S. J. and Hart, O. D. (1983). An analysis of the principal-agent problem, *Econometrica* **51**(1): 7–45.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration, *Journal of Political Economy* **94**(4): 691–719.

Harbring, C. (2007). Why are trust and trustworthiness important for organizations?, *Working Paper*.

Harris, M. and Raviv, A. (1979). Optimal incentive contracts with imperfect information, *Journal of Economic Theory* **20**(2): 231–259.

Hart, O. D. (1995). *Firms, contracts, and financial structure*, Clarendon Press.

Hart, O. D. and Moore, J. (1990). Property rights and the nature of the firm, *Journal of Political Economy* **98**(6): 1119–1158.

Hillier, B. (1997). *The economics of asymmetric information*, MacMillan Press.

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

Holmstrom, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives, *Econometrica* **55**(2): 303–328.

Holmstrom, B. and Tirole, J. (1989). The theory of the firm, *in* R. Schmalensee and R. Willig (eds), *The handbook of industrial organization*, Vol. 1, Elsevier Science, Amsterdam, chapter 2, pp. 61–133.

Ichino, A. and Muehlheusser, G. (2004). How often should you open the door? Optimal monitoring to screen heterogeneous agents, *Journal of Economic Behavior & Organization* **67**(3-4): 820–831.

Ittner, C. D., Larcker, D. F. and Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard, *The Accounting Review* **78**(3): 725–758.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure, *Journal of Financial Economics* **3**(4): 305–360.

Kahneman, D. (2003). A psycholoigical perspective on economcis, *American Economic Review* **93**(2): 162–168.

Kaplan, R. S. and Norton, D. P. (1992). The balanced scorecard: Measures that drive performance, *Harvard Business Review* **70**(1): 72–79.

Klein, B. and Leffler, K. B. (1981). The role of market forces in assuring contractual performance, *Journal of Political Economy* **89**(4): 615–641.

Kurdelbusch, A. (2002). Multinationals and the rise of variable pay in Germany, *European Journal of Industiral Relations* **8**(3): 325–349.

Laffont, J.-J. and Martimort, D. (2002). *The theory of incentives: The principal-agent model*, Princeton University Press.

Lazear, E. P. (1986). Salaries and piece rates, *Journal of Business* **59**(3): 405–431.

Lazear, E. P. (1990). The timing of raises and other payments, *Carnegie-Rochester conference series on public policy* **33**: 14–48.

Lazear, E. P. (2000). Performance pay and productivity, *American Economic Review* **90**(5): 1346–1361.

Lebow, D., Shreiner, L., Slifman, L. and Starr-McCluer, M. (1999). Recent trends in compensation practices, *The Federal Reserve Board, Finance and Economics Discussion Series* .

Levin, J. (2003). Relational incentive contracts, *American Economic Review* **93**(3): 835–857.

Lipe, M. G. and Salterio, S. E. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures, *The Accounting Review* **75**(3): 283–298.

MacLeod, W. B. (2003). Optimal contracting with subjective evaluation, *American Economic Review* **93**(1): 216–40.

MacLeod, W. B. and Malcomson, J. M. (1989). Implicit contracts, incentive compatibility, and involuntary unemployment, *Econometrica* **57**(2): 447–80.

Manoli, D. and Sannikov, Y. (2005). Competitive contracting and employment dynamics, *Working Paper*.

MasCollel, A., Whinston, M. and Green, J. (1995). *Microeconomic theory*, Oxford University Press.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation, *Review of Economic Studies* **38**(114): 175–208.

Mirrlees, J. A. (1974). Notes on welfare economics, information and uncertainty, *in* M. D. Balch, M. and S. Wu (eds), *Essays on economic behavior under uncertainty*, North-Holland, p. 243–258.

Mirrlees, J. A. (1975). The theory of moral hazard and unobservable behaviour: Part 1, *Published 1999 in: Review of Economic Studies* **66**(1): 3–21.

Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization, *Bell Journal of Economics* **7**(1): 105–131.

Mitusch, K. (2006). Non-commitment in performance evaluation and the problem of information distortions, *Journal of Economic Behavior & Organization* **60**(4): 507–525.

Moen, E. R. and Rosén, A. (2005). Performance pay and adverse selection, *Scandinavian Journal of Economics* **107**(2): 279–298.

Moers, F. (2005). Discretion and bias in performance evaluation: The impact of diversity and subjectivity, *Accounting, Organizations and Society* **30**(1): 67–80.

Mullainathan, S. (2002). A memory-based model of bounded rationality, *Quarterly Journal of Economics* **117**(3): 735–774.

Parviainen, S. (2003). Performance-related pay spreads in industry, EIROnline, ID: FI0309202F.

Patterson, G. (1992). Distressed shoppers, disaffected workers promt stores to alter sales commission, *Wall Street Journal*.

Pearce, D. G. and Stacchetti, E. (1998). The interaction of implicit and explicit contracts in repeated agency, *Games and Economic Behavior* **23**(1): 75–96.

Phelan, C. (1995). Repeated moral hazard and one-sided commitment, *Journal of Economic Theory* **66**(2): 448–506.

Piccione, M. and Rubinstein, A. (1997). On the interpretation of decision problems with imperfect recall, *Games and Economic Behavior* **20**(1): 3–24.

Prendergast, C. (1999). The provision of incentives in firms, *Journal of Economic Literature* **37**(1): 7–63.

Prendergast, C. (2000). What trade-off of risk and incentives?, *American Economic Review* **90**(2): 421–425.

Prendergast, C. (2002a). The tenuous trade-off between risk and incentives, *Journal of Political Economy* **110**(5): 1071–1102.

Prendergast, C. (2002b). Uncertainty and incentives, *Journal of Labor Economics, Supplement* **20**(2): S115–S137.

Prendergast, C. and Topel, R. (1993). Discretion and bias in performance evaluation, *European Economic Review* **37**(2-3): 355–65.

Prendergast, C. and Topel, R. (1996). Favoritism in organizations, *Journal of Political Economy* **104**(5): 958–978.

Riley, J. C. (2001). Silver signals: Twenty-five years of screening and signaling, *Journal of Economic Literature* **39**(2): 432– 478.

Ross, S. A. (1973). The economic theory of agency: The principal's problem, *American Economic Review* **63**(2): 134–139.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information, *Quarterly Journal of Economics* **90**(4): 629–649.

Rubinstein, A. (1979). Equilibrium in supergames with the overtaking criterion, *Journal of Economic Theory* **21**(1): 1–9.

Salanié, B. (1997). *The economics of contracts*, The MIT Press.

Sarafidis, Y. (2007). What have you done for me lately? Release of information and strategic manipulation of memories, *The Economic Journal* **117**(518): 307–326.

Schmidt, K. M. and Schnitzer, M. (1995). The interaction of explicit and implicit contracts, *Economics Letters* **48**(2): 193–199.

Schweizer, U. (1999). *Vertragstheorie*, Mohr Siebeck.

Shapiro, C. and Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device, *American Economic Review* **74**(3): 433–444.

Shavell, S. (1979). Risk sharing and incentives in the principal and agent relationship, *Bell Journal of Economics* **10**(1): 55–73.

Simon, H. (1955). A behavioral model of rational choice, *Quarterly Journal of Economics* **69**(1): 99–118.

Spence, M. (1973). Job market signaling, *Quarterly Journal of Economics* **87**(3): 355–374.

Tirole, J. (1999). Incomplete contracts: Where do we stand?, *Econometrica* **67**(4): 741–782.

Under the microscope—staff appraisal frequency on the increase (2005). IRS Employment Review, Online Report.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders, *Journal of Finance* **16**(1): 8–37.

von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*, Princeton University Press.

Wilson, A. (2004). Bounded memory and biases in information processing, *Working Paper*.

Ich versichere, dass ich diese Dissertation selbständig verfasst habe. Bei der Erstellung der Arbeit habe ich mich ausschließlich der angegebenen Hilfsmittel bedient. Die Dissertation ist nicht bereits Gegenstand eines erfolgreich abgeschlossenen Promotions- oder sonstigen Prüfungsverfahrens gewesen.

Dortmund, 21 September 2007

Julia Angerhausen