Essays on Optimal Contracts and Renegotiation.

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Susanne Ohlendorf
aus Bonn.

Bonn 2009

Dekan:	Prof. Dr. Christian Hillgruber
Erstreferent:	Prof. Dr. Urs Schweizer
Zweitreferent:	Prof. Paul Heidhues, PhD

Tag der mündlichen Prüfung: 20.7.2009

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn (http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

Abstract

Contract theory studies the incentives and contractual outcomes in economic interactions, and how they are influenced by given institutions and information structures. On an abstract level, a contractual relationship is characterized by the costs and obstacles that have to be faced in order to carry out the desired economic transaction. Such transaction costs include for example search and information cost, bargaining and contracting costs, and enforcement costs. Depending on the type of transaction costs, optimal contracts will take one form or another. To try to understand economic interactions at this level of detail is of course an enormous undertaking, to which this thesis makes a small contribution, focusing on the effect of renegotiation on the form of contracts.

In the first chapter, we consider a repeated moral hazard problem, where both the principal and the wealth-constrained agent are risk-neutral. In each of two periods, the principal can make an investment and the agent can exert unobservable effort, leading to success or failure. Incentives in the second period act as carrot and stick for the first period, so that effort is higher after a success than after a failure. If renegotiation cannot be prevented, the principal may prefer a project with lower returns; i.e., a project may be "too good" to be financed or, similarly, an agent can be "overqualified."

The second chapter examines the efficiency of expectation damages as a breach remedy in a bilateral trade setting with renegotiation and relationship-specific investment by the buyer and the seller. As demonstrated by Edlin and Reichelstein (1996), no contract that specifies only a fixed quantity and a fixed per-unit price can induce efficient investment if marginal cost is constant and deterministic. We show that this result does not extend to more general payoff functions. If both parties face the risk of breaching, the first best becomes attainable with a simple price-quantity contract.

In the third chapter, we consider the case of an upstream seller who works to improve an asset that has been specialized to a downstream buyer's needs. There is no contract; instead the buyer afterwards makes a take it or leave it offer to the seller. We assume that the seller has private information about the fraction of the surplus that he can realize on his own, and show that this leads to higher investment compared to the complete information case. While a seller with a large default payoff has always strong incentives to invest, now also a seller with a low outside option can choose a large investment, trying to convey the impression of having profitable alternatives. This positive effect on investment is traded off against the occurence of inefficient separation, which

results when the buyer mistakenly tries to call the seller's bluff with a low offer.

The forth chapter studies infinitely repeated two player games with perfect information and side payments. Each period consists of two stages, one in which the parties simultaneously choose an action and one in which they make a monetary transfer. We show that in order to find subgame perfect or renegotiation-proof payoffs for a given discount factor one can restrict the analysis to a class of simple stationary strategies, which we call stationary contracts. We provide simple conditions that characterize renegotiation-proof stationary contracts, and apply these to a series of examples. In particular, we show that in a principal-agent game, in which only the agent chooses an action, all Pareto efficient outcomes can be made renegotiation-proof.

Acknowledgements

I am very thankful for the excellent teachers and research conditions that I have found here at the Bonn Graduate School of Economics. In particular, my gratitude goes to my supervisor Urs Schweizer who expertly guided my research and inspired me with his enthusiasm for law and economics. Thanks also to Paul Heidhues for acting as a referee on my thesis committee and for many helpful comments. I gratefully acknowledge financial support from the German Science Foundation through SFB TR 15 and the BGSE.

This thesis is based on four papers that have been presented at conferences, workshops and summer schools. They have greatly improved in the process of receiving comments from participants, colleagues and referees. An article that is based on Chapter 3 of this dissertation has been published in advance in the American Economic Review (Ohlendorf 2009). Chapters 2 and 4 originate from joint work with Patrick Schmitz, whom I would like to thank for sharing his knowledge and experience with me. It has been a pleasure to collaborate with him, as well as with my colleague Sebastian Kranz, whose creativity makes it always fun to discuss economic questions. Chapter 5 is based on a joint project that resulted from some of these discussions.

I also want to thank all my colleagues and former collegues of the third floor for advice and many discussions at tea or coffee breaks. I would also like to take this opportunity to mention Switgard Feuerstein for her role in my decision to do a PhD in Economics. Without her encouragement this thesis would certainly not exist. The second reason that convinced me to start this endeavor was the open and friendly atmosphere that welcomed me at the graduate school. The friendship I found continued to be a stabilizing factor in the years that I have spent here.

I am grateful to my parents who support me in everything I do, and to my sisters and friends, who provide the important balance to my work. Finally, I would like to thank Bastian Goldlücke, for proofreading pages that have nothing to do with his field of research, and for everything else.

Contents

1

Introduction

1.1 Preliminaries

Many exchanges of goods or services are far more complex than ideal market transactions. Apart from frictions like search and coordination costs or contract writing costs, possible obstacles are that actions of others cannot easily be monitored and controlled, that information may be asymmetrically distributed or that utility may not be perfectly transferable. Even if all concerned parties possess all relevant information, outcomes may be difficult to verify in front of a court, and the legal system may be imperfect or restrict the clauses in a contract for other reasons. Moreover, nothing prevents the parties from tearing up their contract and writing a new one. The possibility of renegotiation is a direct consequence of freedom of contracting, and cannot easily be excluded. Renegotiation interacts with the above mentioned market imperfections, sometimes being a blessing sometimes a curse.

The incentive problems created by unobserved actions or private information are typically dealt with by ex ante committing to ex post inefficient behavior. Consider for example the case that one of the contracting parties could take a costly action that increases the value of the relationship but is not observed by the other party. Examples include an employee's working effort, an insured person's care to avoid an accident or a seller's investment into quality of a good or service. A formal contract can only make use of the relationship between the hidden action and the verifiable value it creates, rewarding good outcomes and punishing bad outcomes. Such a punishment, however, is typically inefficient; it could for example leave too much risk with the more risk-averse party, or lead to underemployment. Therefore, once the action lies in the past, the parties may be tempted to ignore the specifications of the original contract and renegotiate to a better outcome. Since contracts that rely on Pareto dominated threats will not be adhered to, the impossibility to exclude renegotiation acts as a constraint on feasible outcomes.

In a world of complete contracts, in which it is possible to write perfectly enforceable contracts on all variables that are observable, there are no allocations that can be reached by renegotiation only. Even if the contracting parties do not have all relevant information at one point in time, they can condition the contract on all events that could possibly arise at later dates. Hence, under the strong assumption of complete contracting, there are no benefits to renegotiation, only potential costs.[1]

Since the ability to commit is valuable, we may observe behavior that seems irrational at first but can be explained by the search for a commitment device, i.e., we may observe agents who deliberately restrict their options, like Odysseus tying himself to the mast to escape the sirens. A principal (an employer or a lender) may choose an inefficient monitoring technology, which she would never do if renegotiation could be excluded, because not learning the real quality of the agent (the employee or borrower) allows her not to be too forgiving of failures. That imperfect observability may help to overcome a commitment problem in environments with incomplete information is shown in Cremer (1995) and Dewatripont and Maskin (1995a). Chapter 2 provides a new example of this effect in a hidden action framework, in which a worse agent is preferred over a more productive one, as the principal gains the credibility to fire the agent.

In contrast to the complete contracting environment, renegotiation can play a beneficial role if contracts are incomplete. If there are many different events that could possibly occur, each with small probability, a contract that provides for all these events may be very costly to write. These costs could be saved by writing only a simple contract and renegotiating it later when circumstances have become more clear. There is, however, a tradeoff to retaining flexibility until the optimal contractual terms present themselves, since outcomes that are desirable ex post do not have to be desirable from an ex ante point of view. If the share of the gains from trade that each of the trading partners receives is determined at a later stage, incentives to prepare for trade depend on the ex post bargaining process. In particular, parties will generally be very reluctant to invest into the relationship in the absence of a binding contract, as the rents that the investment generates have to be shared with the other party according to bargaining power (such underinvestment problems are known by the term *hold-up* problem, see Grout (1984), Klein, Crawford

[1] See for example Fudenberg and Tirole (1990), who explicitly introduce a "renegotiation proofness constraint" in a dynamic contracting problem. For an overview of the effects of such a constraint, see Dewatripont and Maskin (1990) or Bolton and Detwatripont (2005, Chapter 9).

and Alchian (1978), Hart and Moore (1988)). In contrast, renegotiation may not be an obstacle for efficient ex ante investment if at least some contracts can be formulated and enforced. Even a simple contract may have powerful effects on the parties' bargaining positions, and therefore influence the overall outcome. Articles like Hermalin and Katz (1991), Edlin and Reichelstein (1996) and Evans (2008) show that in such cases it may be possible to have ex post flexibility and ex ante efficiency at the same time.

Another reason why contracts may have to be simple is that not all clauses are enforceable. For example, while a contract can specify fines for nonperformance ("liquidated damages"), there is a limit to these fines. In many legal systems these are not allowed to be grossly higher than the expected loss of breach. Uncertainty with regard to which clauses are allowed may lead to an unmerited prevalence of standard remedies. Therefore, standard remedies for breach of contract are an important object of study, even though it is in principle possible to contract around them whenever they are not suitable. In Anglo-American law, the default legal remedy for breach of contract is the expectation damages rule. This rule, and how it can help the contracting parties to achieve maximal gains from trade in a framework with simple noncontingent contracts and renegotiation, is the topic of Chapter 3.

The theory of incomplete contracts is closely linked to the theory of the firm (see for example Coase (1937) and Williamson (1971, 1979, 1985)). This theory offers a way to explain how production is organized: which transactions are conducted within a firm and which via a market. The idea is that because contracts are never perfect, there is a difference between doing something yourself and writing a contract in order to get somebody else to do it. Williamson identifies three major determinants of this "make or buy" decision: asset specificity, uncertainty, and frequency of transactions. The impossibility of writing complete contracts and the resulting potential for holdup if investment is specific to the relationship encourage contracting parties to enter long-term relationships or to vertically integrate.

In the property rights models that study these decisions (Grossman and Hart (1986), Hart and Moore (1990), Hart (1995)) only ownership structures can be contracted on. Outside opportunities, which depend on who owns the asset, are then crucial to determine the amount of investment. Firms that can put their assets to profitable alternative uses are less vulnerable to holdup and therefore can invest more. In Chapter 4, we study the possibility that asymmetric information about the outside option could lead to higher investment on average, as firms try to signal a high outside option with their investment.[2]

Another way to overcome enforcement problems and induce more investment is a long-term relationship. If the relevant actions are observable to the contracting parties, but it is not possible to write a contract that has to be enforced by a third party, then reputation and repeated interaction may be used instead. Informal agreements have to be enforced within the relationship, e.g., by threats of terminating a profitable relationship. The resulting agreements are not contracts in the legal sense, but are still sometimes called *relational contracts*. The term originated in the legal literature, in work by Macaulay (1963) and MacNeil (1978). The study by Macaulay documents how firms in practice coordinate their behavior without the assistance of written contracts. Although detailed agreements are made use of, disputes are often settled without reference to the contract.

When economists speak of relational contracts, they mean "informal agreements sustained by the value of future relationships"(Baker, Gibbons and Murphy (2002, p.39)). Concern for the future of the relationship is modeled by letting the same stage game be repeated infinitely often. Relational contracts are essentially subgame perfect equilibria in such a repeated game, but the term also implies a kind of relationship between the parties.[3] This relationship may give them the opportunity to negotiate to efficient equilibria, and also to organize monetary transfers to induce compliance with jointly optimal behavior. In Chapter 5, we assume that parties can make unlimited monetary transfers and impose a renegotiation-proofness constraint, and characterize the resulting outcomes for a general class of games.

In the remainder of this introduction we summarize the chapters and in each summary highlight the imperfections in the economic environment that the chapter focuses on. A detailed introduction including an overview of related literature is given at the beginning of the chapters.

---

[2] Signaling may occur if some information may be conveyed through observable actions prior to the contracting. The classical example is education as a wasteful signal for productivity (Spence (1973)).

[3] This could for example be a business to business relationship, or an employer - employee relationship within a firm. Baker et al. (2002) entertain the idea that the role of managers is to communicate and adjust the relational contracts within a firm. See also McLeod (2007) for a discussion of how to define relational contracts.

1.2 Optimal renegotiation-proof contracts with dynamic investment

The second chapter studies a problem in classical principal-agent theory.[4] This theory is concerned with the relationship between a principal (an employer, a landlord, a bank or other investor etc.) and an agent who acts on behalf of the principal. The problem of the principal is to design a compensation scheme that aligns the interests of the agent as best as possible with his own, taking into account the unobservability of the agent's actions and other technological constraints.

In the chapter, we highlight the effect of renegotiation in a repeated hidden action model featuring a principal who delegates a project to a wealth-constraint agent. Since effort is not contractible, the agent must be motivated with a bonus in case of success or even by a threat to terminate the project after a failure in the first period. Although termination is an inefficient second period outcome, such a contract may ex ante be optimal for the principal. If renegotiation cannot be prevented, a threat to terminate ceases to be credible for projects with high potential. Only projects that yield a low per-period return will actually be discontinued, and such a project then could be preferred by the principal. The worse project has a value as a commitment device, and this effect on effort may actually outweigh the smaller return.

The chapter makes two contributions to the literature. First, it shows how effort levels compare across states in a two-period moral hazard model with limited liability. The principal exploits the repetition by rewarding two successes in a row relatively more than the individual successes. This leads to what we call the *hot-hand* effect: an agent who achieves a success in the first period is better motivated and hence more likely to succeed in the second period as well. The second contribution is to show that if renegotiation is introduced in this repeated version of a standard model, an *overqualification* effect may occur: it can happen that an optimizing principal chooses a worse project/agent over a better one.

In the environment studied here, lack of observability and non-transferability of utility makes the first best unattainable. While these problems are more severe in the one-shot interaction, they can only in part be overcome by the two-period contract. The result is generated by the presence of both these imperfections: if the agent was not wealth constrained, he could simply buy the firm and become residual claimant, while if his effort choice was observable, he could be forced to exert first best effort. Lack of commitment plays a role because punishment in form of a low-powered incen-

[4] This chapter is based on Ohlendorf and Schmitz (2008), available as a CEPR discussion paper.

tive contract -or even termination after a failure- is Pareto dominated by the optimal one-period contract.

There are many versions of the renegotiation-proofness principle[5], but in this complete contracting environment, in which the principal optimizes among all possible contracts, it takes a very simple form: the optimization can be restricted to contracts that specify Pareto optimal outcomes in every subgame. This means that the induced second period effort must be weakly greater than the second best effort of the one-shot problem, and that termination can only occur if the project generates a sufficiently low return. The resulting contract is never renegotiated, and thus would not change if the ex post bargaining process was different from the situation ex ante, in which the principal has all the bargaining power.

1.3 Simple efficient contracts and contract law

Contract enforcement not only guarantees that payments are made and delivery takes place, but may also foster investments which are of value only in a particular relationship. If buyers can rely on a contract, they can plan an advertisement campaign or start training their workers to use contracted-for equipment before actual delivery. Likewise, sellers can engage in research to reduce production costs and tailor their production to the needs of the buyer without fear of holdup. Whether a contract indeed leads to the right incentives for investment also depends on the consequences of breach. If a contract is breached, damages have to be paid, which are either stipulated in the contract, or set by a court. Chapter 3 is concerned with the most common legal rule for damage payments, the expectation measure.

The chapter builds on the analysis of Edlin and Reichelstein (1996), who analyze standard breach remedies when contracts consist only of an up front transfer, a quantity to be traded and a price per unit. Because contracts can be renegotiated once state uncertainty has been resolved, the ex post optimal decision is always implemented. The up-front transfer can be used to divide the gains from trade, such that the role of price and quantity is merely to induce ex ante efficient investment. Edlin and Reichelstein show that these variables may successfully fill this role if only one of the contracting parties invests. The bilateral investment case is more difficult, but for specific performance it is sometimes possible to align the incentives of both parties with a single quantity. Expectation damages, on the other hand, seems ill-suited for bilateral investment cases.

---

[5] See Brennan and Watson (2002).

The intuition for the inefficiency of expectation damages is that this remedy treats the breaching party and the victim of breach asymmetrically. While the breaching party takes an efficient breach decision, and in anticipation of breach will have efficient investment incentives, the victim of breach is always insured to receive the expected profit from the contracted quantity. This party then prepares for a higher than optimal quantity, not internalizing the risk of breach. The contribution of Chapter 3 of this thesis is to recognize that in the framework of Edlin and Reichelstein, with divisible contracts, the identity of the breaching party is endogenous and influenced by the per-unit price. It is shown that if one moves away from the deterministic and linear functions, both parties face the risk of breaching. The two variables price and quantity can be used to fine-tune the investment incentives of the two parties, balancing the hold-up effect that arises from the possibility of renegotiation against the overinvestment effect that arises from being the victim of breach.

This result holds only when contracts are divisible and payoff functions are sufficiently concave. If these conditions do not hold and a simple price-quantity contract cannot reach the first best, one might ask how comprehensive contracts have to be for an efficient outcome under the legal regime of expectation damages. It is demonstrated in the chapter how a slightly more complicated contract with a stochastic price can reach the first best for a more general class of functions, including the linear ones. An option contract may also help in case of a deterministic and linear cost function, but not in general, since whenever only one party faces the risk of breaching, Edlin and Reichelstein's inefficiency result prevails.

In this contracting environment, the optimal trade decision becomes contractible ex post but cannot be described ex ante. If the parties were to go to court, the court would be able to assess damages, but parties do not make use of this information directly in the contract. This can be viewed as an adhoc restriction of possible contracts, typical for the law and economics literature, which often assumes simple contracts in order to study the effect of a given institution like a breach remedy. One may alternatively take institutions as given and study how complex contracts have to be in reaction to the prescribed rules. In the special case of Chapter 3 with costless renegotiation, the parties are able to achieve an efficient outcome with a simple contract. Assuming small contract writing costs, such a non-contingent contract should then be strictly preferred to more complex agreements.[6] The optimal contract looks natural in the sense that the stipulated prices and quantities do

[6] Of course, the model abstracts from a lot of other frictions like legal costs, bargaining costs etc. as well. The effects of introducing small costs for different actions certainly deserve more scrutiny, as it is not always clear how these costs would interact with each other.

not take on extreme values, but the exact values depend on the details of the environment, in particular on the anticipated bargaining process.

1.4 Incomplete contracts and asymmetric information about asset specificity

In many industries big companies rely on the relationship-specific investments of their suppliers, yet the subcontractors are small firms compared to their customers, and potential customers are few. If the bargaining power lies entirely with the customer, how are the necessary innovations induced in this environment where the customer dictates the rules and sueing for payment is unthinkable? It seems vital for small suppliers to have many customers such that in case of separation they can make up for the loss by dealing with others. Suppliers who manage not to be dependent on any one customer may be able to avoid exploitation and be compensated for their investment.

In Chapter 4, we show that information rents resulting from asymmetric information about the position of the supplier in the market may stimulate innovation if the customer has no other way to commit herself to adequately reward investment. The idea is that too much pressure on the buyer's subcontractors would result in the loss of the strong ones, while only the weak ones would stay, hurting also the buyer itself. Moreover, there is a signaling motive in the investment choice. If the best alternative use of the relationship-specific asset is private information to the supplier, the customer will try to deduce the outside option from the level of investment. If the supplier is very reluctant to invest, it may be from fear of hold-up because of a low outside option, and the buyer will make only a low offer. If instead the supplier is very eager to invest, it seems likely that the private value from the investment is high, hence the buyer has to make a high offer. Now, of course, the possibility arises that a supplier with a low outside option mimics the type with the high outside option and invests more. This effect may mitigate the hold-up problem and lead to higher investment.

We find that while the information asymmetry leads to higher investment, this effect is traded against the inefficiency generated by the non-investing party trying to appropriate part of the information rents. Welfare comparisons can therefore go in both directions. As in Chapter 3, the parties are interacting in an incomplete contracting environment. A change in bargaining power would lead to a different equilibrium outcome. In particular, giving all bargaing power to the investing party would lead to an efficient allocation. As in Chapter 2, limited commitment also plays a role in making the outcome more inefficient and less profitable for the buyer.

1.5 Renegotiationproof relational contracts

The last chapter studies renegotiation-proof relational contracts.[7] There are two parties who find themselves in the same situation in a potentially infinite number of periods. They can be business partners who trade repeatedly or cooperate in production, or firms in the same market who compete for the same pool of workers or collude in price setting. Each period consists of two stages, one stage in which the parties simultaneously choose an action and one stage in which they can make a payment to the other party. The parties know the entire history of the game, but compliance with the relational contract cannot be enforced by verifiable contracts.

Folk theorems tell us that as the parties become infinitely patient, all individually rational outcomes can be sustained as subgame perfect equilibria in the repeated game. The future matters to such a great extent that the players are afraid to deviate from any agreement if a devation leads to a worse outcome. The two parties might thus be trapped in a very bad equilibrium, but such outcomes seem unrealistic if they are able to communciate with each other. In case communication is feasible, equilibria that can never be renegotiated to better outcomes seem more likely than others. This idea is captured formally by requiring equilibria to be renegotiation-proof. Since renegotiation-proofness is not straightforward to define for infinitely repeated games, there exist several definitions in the literature. In the chapter, we adapt strong perfection as defined in Rubinstein (1980) as well as Farell and Maskin (1989)'s concepts of weak and strong renegotiation-proof equilibria to our setting with monetary transfers. We assume that bargaining can take place after every period and also between the two stages of a given period, and provide a characterization of renegotiation-proof relational contracts given arbitrary discount factors.

We find that all Pareto optimal subgame perfect payoffs and renegotiation-proof payoffs can be found by restricting attention to a class of stationary contracts. These stationary contracts prescribe play of the same action in every period on the equilibrium path, and in case of a deviation allow the deviator to pay a fine and return to equilibrium play. The actual punishment that results if a fine is not paid occurs within one period. Fines can hence be used to create one-period punishments, which are more resistent to renegotiation than a punishment that lasts forever. If a customer has not paid yet, instead of never trading with him again, it may be more credible to threaten no delivery in this month only and expect a better contract for the next month. Renegotiations in this month are then blocked by the plans for the next. Not

[7] The chapter is based on Kranz and Ohlendorf (2009), available as a SFB TR 15 discussion paper.

paying and trying to renegotiate will be too high a temptation only if a one month delivery stop is too costly.

The concepts of renegotiation proofness that are applied in the chapter define properties that an equilibrium has to fulfill in order to be considered renegotiation-proof. Behind this axiomatic formulation lies the idea that once there was a deviation, there is time for bargaining before the next action has to be chosen. The concepts differ with respect to the timing of negotiations and the set of alternative strategies that the parties bargain about, but they all have in common that the default option should the negotiations fail is the original equilibrium. In addition, they do not specify an explicit bargaining process and therefore usually do not make a unique prediction. Bargaining in the beginning of the game then is different from bargaining at later stages.

2

Repeated moral hazard, limited liability, and renegotiation

2.1 Introduction

Consider a risk-neutral principal, who can invest to install a project and hire a risk-neutral but wealth-constrained agent. The agent can exert unobservable effort, which increases the likelihood of success. In the one-shot problem, there is a well-known trade-off between effort incentives and rent extraction, which leads to a downward distorted effort level compared to the first-best solution. We extend the standard model by assuming that there is a second period, in which the principal can again make an investment to continue the project and the agent can again exert unobservable effort. It turns out that there are several interesting insights that so far have escaped the literature on repeated moral hazard, which was focused on the case of risk-averse agents.

In particular, if the principal can commit not to renegotiate, the second period incentives can be used to partially circumvent the limited liability constraint. In the second period, the principal implements a particularly high effort level following a first-period success and a particularly low effort level following a first-period failure. The prospect of a higher second-period rent following a first-period success motivates the agent to exert more effort in the first period; i.e., rents in the second period act as reward and punishment for the first period. It should be emphasized that we assume no impact of a first-period success or failure on the second-period technology. Nevertheless, if an outsider observed today a principal-agent pair that was successful and another identical pair that was not successful, he would be right to predict that the first pair also is more likely to succeed tomorrow. In other words, a "hot hand" effect is generated endogenously, merely based on incentive considerations. [1]

[1] The term "having a hot hand" originated in basketball and means having a streak of successes that cannot be attributed to normal variation in performance. It

Just as in the one-shot model, effort levels are distorted and not every project that would be installed in a first-best world will be pursued under moral hazard. It also is still the case that the principal will always prefer a project (or, equivalently, an agent) that can yield a larger return (among otherwise identical projects or agents). Somewhat surprisingly, however, the latter observation is no longer true if renegotiation cannot be ruled out.

The "hot hand" effect implies that a principal would sometimes like to commit to terminate a project following a first-period failure, even though technologically the success probability of the second period is not affected by the first-period outcome. Yet, the threat to terminate may not be credible if renegotiation cannot be prevented. In this case, a new kind of inefficiency occurs: The principal might deliberately choose a project that is commonly known to yield smaller potential returns than another (otherwise identical) project that is also available. Similarly, she might deliberately hire an agent that is commonly known to be less qualified.

The reason that a project might be "too good" to be funded or an agent might be "overqualified" is the fact that the principal cannot resist the temptation to renegotiate if the potential return is too attractive, which is anticipated by the agent, whose incentives to work hard in the first period are dulled. In contrast, a less qualified agent or an agent working on a less attractive project may well be willing to exert more effort, because he knows that in case of a failure he will not get a second chance. Since the credible threat to terminate the project after a first-period failure improves first-period incentives, there are indeed parameter constellations under which a relatively bad project is funded, while a better project is not.

The literature on repeated moral hazard problems and renegotiation has different strands. Most papers consider repeated versions of the traditional moral hazard setting, where the agent is risk-averse and there is a trade-off between insurance and incentives.[2] In a pioneering paper, Rogerson (1985) considered a two-period moral hazard problem and showed that the optimal second-period incentives depend on the first-period outcome, even though the periods are technologically independent. However, his result is driven by the consumption-smoothing motive of the risk-averse agent,[3] which is absent in our setting. In moral hazard models with a risk-averse agent, renegotiation

seems to spectators that the probability of a success increases after a row of successes, even though the trials in question are independent; see Gilovich, Vallone, and Tversky (1985).

[2] For comprehensive surveys, see Chiappori et al. (1994) and Bolton and Dewatripont (2005, ch. 10).

[3] Cf. Malcomson and Spinnewyn (1988), Fudenberg, Holmström, and Milgrom (1990), and Rey and Salanié (1990).

is an issue even in the one-shot problem, because after the agent has chosen an effort level, there is no need to expose him to further risk. Fudenberg and Tirole (1990), Ma ((1991), (1994)) and Matthews ((1995), (2001)) show that it depends on the details of the renegotiation game whether or not effort incentives are reduced.[4] In contrast, in our framework there is scope for renegotiation only if the moral hazard problem is repeated, and the details of the renegotiation game are irrelevant for our results.

Although we consider a repeated moral hazard problem, it is interesting to note that our results are also related to the repeated adverse selection literature.[5] Specifically, in a seminal paper Dewatripont and Maskin (1995b) consider a two-period model where the agent has private information about the quality of a project that he submits for funding. Ex ante, the principal would like to terminate bad projects after the first period in order to deter the agent from submitting them ("hard budget constraint"). Yet, at the beginning of the second period she is tempted to refinance them ("soft budget constraint"). The absence of commitment power thus enables bad projects to be funded. However, as has been pointed out by Kornai, Maskin, and Roland (2003, p. 1110), the principal would not finance a bad project if she knew the quality ex ante. In contrast, in our model a bad project may be funded, while a better project may not be funded, even though the quality is common knowledge.

The remainder of the chapter is organized as follows. In Section 2.2.1, we introduce the one-shot moral hazard problem with a risk-neutral but wealth-constrained agent, which now is sometimes called "efficiency wage" model.[6] This model serves as a benchmark for the dynamic analysis. We then introduce the two-period model in Section 2.2.2.[7] In Section 2.3, we analyze the

---

[4] See also Hermalin and Katz (1991) and Dewatripont et al. (2003), who consider observable but unverifiable effort.

[5] The fact that the one-shot moral hazard model with a risk-neutral but wealth-constrained agent has some similarities to the one-shot adverse selection model has already been noted by Laffont and Martimort ((2002),p. 147).

[6] See Tirole (1999, p. 745) or Laffont and Martimort (2002, p. 174). See also Innes (1990), Pitchford (1998), and Tirole (2001) for more detailed discussions of the one-shot moral hazard model with risk-neutrality and resource constraints. Moreover, cf. the traditional efficiency wage literature (Shapiro and Stiglitz, (1984)) and the literature on deferred compensation (Lazear, (1981); Akerlof and Katz, (1989)), which are related but have a different focus. In related frameworks, Strausz (2006) studies auditing and Lewis and Sappington (2000) explore the role of private information about limited wealth.

[7] Dynamic models with risk-neutral agents, hidden actions, and wealth constraints include also Crémer (1995), Baliga and Sjöström (1998), Che and Yoo (2001), and Schmitz (2005). Yet, they rely on features (private information about productiv-

commitment scenario and highlight the "hot hand" effect. In Section 2.4, it is assumed that renegotiation cannot be ruled out, which leads to the "overqualification " effect. Finally, concluding remarks follow in Section 2.5. All proofs have been relegated to the end of the chapter.

2.2 The model

2.2.1 The one-shot contracting problem

As a useful benchmark, let us first take a brief look at the one-shot moral-hazard problem that will be repeated twice in our full-fledged model. There are two parties, a principal and an agent, both of whom are risk-neutral. The agent has no resources of his own, so that all payments to the agent have to be nonnegative. The parties' reservation utilities are given by zero. At some initial date 0, the principal can decide whether or not to pursue a project. If she installs the project, she incurs costs I and she offers a contract to the agent. Having accepted the contract, the agent exerts unobservable effort $e \in [0, 1]$ at date 1. His disutility from exerting effort is given by $c(e)$. Finally, at date 2, either a success ($y = 1$) or a failure ($y = 0$) is realized, where $\Pr\{y = 1|e\} = e$. The principal's verifiable return is given by yR.

ASSUMPTION 1 *The effort cost function satisfies*[8]

1. $c \in \mathcal{C}^3([0, 1])$ *and* $c'(1) > R$,
2. $c' > 0$, $c'' > 0$, $c''' \geq 0$,
3. $c(0) = 0$ *and* $c'(0) = 0$.

The first-best effort level maximizes the expected total surplus $S(e) := eR - c(e)$ and is characterized by

$$S'(e^{FB}) = R - c'(e^{FB}) = 0.$$

In a first-best world, the project would be installed whenever $S(e^{FB}) \geq I$.

The principal could attain the first-best outcome, but in order to do so she would have to leave all of her returns to the agent, because payments to the agent cannot be negative. Hence, the principal faces a trade-off between

ity, observable yet unverifiable effort, common shocks, and technological relations between the periods, respectively) which are absent in the repeated (pure) moral hazard problem studied here.

[8] The assumptions allow us to focus on first-order conditions. With minor notational modifications of the proofs, we could relax Assumption 1a) by including $c \in \mathcal{C}^3([0, 1))$, where $\lim_{e \to 1} c'(e) > R$.

increasing the pie and getting a larger share for herself. To find the second-best solution, observe first that the principal will not pay anything when no revenue is generated. Next, let t denote the principal's transfer payment to the agent in case of success. The agent's expected payoff from exerting effort e is $et - c(e)$. If $t \leq R$, which obviously is optimal, the agent's maximization problem has an interior solution characterized by $t = c'(e)$. Hence, the principal maximizes $P(e) := e[R - c'(e)]$ over e. The first-order condition characterizing the effort level she will implement thus is

$$P'(e^{SB}) = R - c'(e^{SB}) - e^{SB}c''(e^{SB}) = 0.$$

Note that the function P is concave, positive for $e < e^{FB}$ and negative for $e > e^{FB}$. We also define $A(e) := ec'(e) - c(e)$, the agent's rent from a contract that leads him to choose effort e. Since its derivative is $A'(e) = ec''(e)$, this is a non-negative and strictly increasing function. Hence, a higher implemented effort level yields higher rents for the agent. In order to reduce the agent's rent, the principal thus introduces a downward distortion of the induced effort level, $e^{SB} < e^{FB}$.

In the one-shot problem, the principal will install the project whenever $P(e^{SB}) \geq I$; i.e., not all projects that would be pursued in a first-best world will actually be installed. However, given the choice between two (otherwise identical) projects with possible returns $R = R_g$ and $R = R_b < R_g$, the principal will never prefer the bad project that can yield R_b only.

2.2.2 The two-period model

Now we turn to the full-fledged two-period model. For simplicity, we neglect discounting. At date 0, the principal decides whether ($x_1 = 1$) or not ($x_1 = 0$) to invest an amount $I_1 > 0$ in order to install the project and run it for the first period. If she pursues the project, she makes a take-it-or-leave-it contract offer to the agent. Having accepted the offer, at date 1 the agent chooses an unobservable first-period effort level $e_1 \in [0, 1]$, incurring disutility $c(e_1)$. At date 2, the verifiable first-period return $y_1 R$ is realized, where $y_1 \in \{0, 1\}$ denotes failure or success, and $\Pr\{y_1 = 1|e_1\} = e_1$. The project may then be terminated ($x_2(y_1) = 0$) or continued ($x_2(y_1) = 1$) which is verifiable.[9] In

[9] We assume that it is too costly for the principal to replace the agent at date 2, because at that point in time the parties are "locked-in" (i.e., the relationship has undergone Williamson's (1985) "fundamental transformation"). For instance, hiring a new agent for the ongoing project might require specific training, which makes replacement unprofitable. See Spear and Wang (2005) and Mylovanov and Schmitz (2008) for models in which replacement involves no costs. Of course, our model could be extended to the case of costly replacement, but this would make the exposition less tractable.

order to continue the project, the principal must invest an amount $I_2 \leq I_1$.[10] In this case, at date 3 the agent chooses an unobservable second-period effort level $e_2(y_1) \in [0,1]$. Finally, at date 4 the verifiable second-period return $y_2 R$ is realized, where $y_2 \in \{0,1\}$ and $\Pr\{y_2 = 1 | e_2(y_1)\} = e_2(y_1)$, and the contractually specified transfer payments are made. Note that the two periods are independent; in particular, we do not assume any technological spillovers that would make a second-period success more likely after a first-period success.

The sequence of events is illustrated in Figure 2.1.

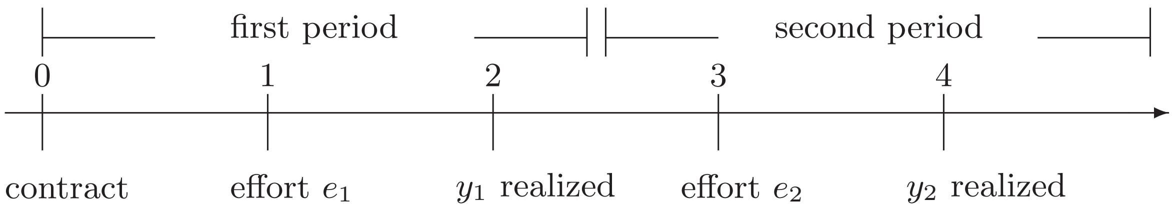

Fig. 2.1. Repeated moral hazard problem.

The first-best solution. Assume for a moment that effort was verifiable. If $S(e^{FB}) + S(e^{FB}) \geq I_1 + I_2$, then the principal would install the project ($x_1 = 1$), she would continue regardless of the first-period outcome ($x_2(0) = x_2(1) = 1$), and she would implement $e_1 = e_2(0) = e_2(1) = e^{FB}$ with a straightforward forcing contract, leaving no rent to the agent. Otherwise, the project would not be installed at all.

Contracts when effort is unobservable. In the remainder of the paper, we assume again that effort levels are unobservable. If the principal decides to install the project ($x_1 = 1$), a contract can specify the continuation decision $x_2(y_1)$ and transfer payments $t(y_1, y_2) \geq 0$ from the principal to the agent. Note that we do not impose any ad hoc restrictions on the class of feasible contracts; i.e., there is complete contracting in the sense of Tirole (1999).[11] It

[10] The assumption $I_2 \leq I_1$ is made only to simplify the presentation. If $I_2 > I_1$, then the project might be terminated even after a success in the first period, which would lead to more case distinctions without yielding any additional economic insights.

[11] As is well-known, in a pure moral hazard framework with unobservable effort, there is no need to consider messages sent by the agent (in contrast, such messages would have to be considered if effort were observable yet unverifiable). Moreover, note that there is no need to consider additional payments made at earlier dates. The principal is (weakly) better off when she makes all payments to the agent at

will be helpful to distinguish between (direct) rewards regarding the first and second period, so we use the following notational convention:

$$t(0,0) = t_0 + t_{00}, \quad t(0,1) = t_0 + t_{01},$$
$$t(1,0) = t_1 + t_{10}, \quad t(1,1) = t_1 + t_{11}.$$

It is easy to see that it is optimal for the principal to set $t_0 = t_{00} = t_{10} = 0$.[12] A first-period success will thus be directly rewarded with a bonus payment t_1, while a second-period success will be rewarded with a bonus t_{01} (following a first-period failure) or t_{11} (following a first-period success). As we will see, a first-period success will also be indirectly rewarded by the prospect of getting a larger bonus for a second-period success if it follows a first-period success, which will be a driving force behind our main results.

2.3 Commitment: The “hot hand” effect

Let us first assume that the principal can commit not to renegotiate the contract that is written at date 0. If the principal installs the project ($x_1 = 1$), her expected profit is given by

$$\begin{aligned} &\Pi(e_1, e_2(0), e_2(1)) \\ &= e_1 \left(R - t_1 + x_2(1)\left[e_2(1)(R - t_{11}) - I_2\right]\right) \\ &\quad +(1 - e_1)x_2(0)\left[e_2(0)(R - t_{01}) - I_2\right] - I_1, \end{aligned}$$

otherwise it is given by zero. The agent’s incentive compatibility constraints for the second period read

$$e_2(0) = \arg\max_e e t_{01} - c(e),$$
$$e_2(1) = \arg\max_e e t_{11} - c(e).$$

The agent’s first-period incentive compatibility constraint is given by

$$e_1 = \arg\max_e e\left[t_1 + e_2(1)t_{11} - c(e_2(1))\right] + (1 - e)\left[e_2(0)t_{01} - c(e_2(0))\right] - c(e).$$

date 4 (since the agent might “consume” earlier payments, so that he cannot be forced to pay them back).

[12] Obviously, the principal will set $t(0,0) = 0$. Hence, we can write $t_0 = t_{00} = 0$ and $t(0,1) = t_{01}$ without loss of generality. Moreover, note that we can restrict attention to $t_{11} \geq t_{10}$. If t_{10} were not zero, the payments could be replaced by $\tilde{t}_{10} = 0$, $\tilde{t}_1 = t_1 + t_{10}$ $\tilde{t}_{11} = t_{11} - t_{10}$. Notice that we could also assume that the direct rewards for the first period are paid at date 2 already.

Moreover, the payments must be non-negative due to the agent's wealth constraint, which together with incentive compatibility also ensures participation.

The following proposition characterizes the second-best solution of the two-period model under full commitment.[13]

Proposition 2.1. *Assume that the principal can commit not to renegotiate. There exist a cut-off level I_2^* and continuous and decreasing threshold functions $I_1^C(I_2)$ and $I_1^T(I_2)$, where $0 < I_2^* < I_1^C(I_2^*) = I_1^T(I_2^*)$.*

a) If $I_2 \leq I_2^*$ and $I_1 \leq I_1^C(I_2)$, then the project is installed and it is always continued, $x_1 = x_2(0) = x_2(1) = 1$. The effort levels induced by the principal's optimal contract satisfy

$$e^{FB} \geq e_2^C(1) > e_1^C > e^{SB} > e_2^C(0) > 0.$$

b) If $I_2 > I_2^*$ and $I_1 \leq I_1^T(I_2)$, then the project is installed and it is terminated whenever the first period was a failure, $x_1 = x_2(1) = 1$, $x_2(0) = 0$. In this case,

$$e^{FB} \geq e_2^T(1) > e_1^T > e^{SB}.$$

c) Otherwise, the project is not installed at all, $x_1 = 0$.

Proof. See section 2.6 at the end of the chapter.

There are two ways in which a success in the first period is indirectly rewarded by the principal. First, consider the case in which the principal's investment costs I_1 and I_2 are sufficiently small, so that the project is installed and always continued (Proposition 1a). Even though a success in the first period has no technological effect whatsoever on the likelihood of a success in the second period, the principal implements $e_2^C(1) > e^{SB} > e_2^C(0)$. Giving the agent in the second period articularly high incentives following a first-period success (and particularly low incentives following a failure) has desirable spillover effects on the first-period incentives: The agent works hard in the first period not only in order to get the direct reward t_1, but also in order to enjoy a higher second-period rent ($A(e_2^C(1))$ instead of $A(e_2^C(0))$). Interestingly, from an outsider's perspective, this means that a success in the second period is indeed more likely to be observed after a first-period success; i.e., there is a "hot hand" effect endogenously generated purely due to incentive considerations.[14] In fact, the direct first-period reward t_1 will be positive

[13] The superscript "C" denotes continuation, while "T" denotes termination.

[14] See also McFall et al. ((2006)), who show a "hot hand" effect in a series of contests when there happens to be a huge reward for those who win the most contests. Players with initial good luck have more to win later on, therefore they exert relatively higher effort.

only if the principal already induces $e_2^C(1) = e^{FB}$, so that implementing an even higher effort level following a first-period success would reduce the total surplus. Since giving the agent incentives in the first period is now cheaper than in the one-shot problem, the principal implements $e_1^C > e^{SB}$.

Second, the fact that the principal induces only low second-period effort after a first-period failure implies that continuing the project after a first-period failure might not be in the principal's interest when her continuation costs I_2 are sufficiently large (Proposition 1b).[15] Clearly, if $e_2^C(0)$ is so small that $P(e_2^C(0)) < I_2$, the principal is better off if she terminates the project. But even if this inequality does not hold, it can still be optimal for the principal to commit to terminate the project, because doing so improves the agent's first-period incentives. Hence, the cut-off level I_2^* is smaller than $P(e_2^C(0))$.

The inefficiencies exhibited by the second-best solution are of a similar nature as the inefficiencies we encountered in the one-shot model. There are downward distortions of the effort levels compared with the first-best solution, and as a result there are projects that would be installed (and continued) in a first-best world, but that are not pursued (or at least not continued after a first-period failure) in the presence of moral hazard. However, it is still impossible for a project to be "too good" to be pursued, as is stated in the following corollary.

Corollary 2.2. *Assume that the principal can commit not to renegotiate. If at date 0 the principal can choose between two (otherwise identical) projects with possible returns $R = R_g$ and $R = R_b < R_g$, she will never prefer the bad project that can yield R_b only.*

Proof. See section 2.6 at the end of the chapter.

2.4 Renegotiation: The "overqualification" effect

After the first period is over, the principal might want to modify the contractual arrangements, because at that point in time she would be best off under the optimal one-period contract as characterized in Section 2.1. In the following we assume that the principal cannot ex ante commit not to renegotiate the contract.[16] In our complete contracting framework, the principal

---

[15] Obviously, given that the principal installed the project at all, she will always continue if the first period was successful, because $I_2 \leq I_1$.

[16] See Bolton and Dewatripont (2005) for extensive discussions of the assumption that renegotiation cannot be ruled out.

can mimic the outcome of renegotiations in her original contract; i.e., we can confine our attention to renegotiation-proof contracts.[17]

Proposition 2.3. *Assume that the principal cannot commit not to renegotiate. There exist continuous and decreasing threshold functions* $\bar{I}_1^C(I_2)$ *and* $I_1^T(I_2)$*, where* $P(e^{SB}) < \bar{I}_1^C(P(e^{SB})) < I_1^T(P(e^{SB}))$*.*

a) If $I_2 \leq P(e^{SB})$ *and* $I_1 \leq \bar{I}_1^C(I_2)$*, then the project is installed and it is always continued,* $x_1 = x_2(0) = x_2(1) = 1$*. The effort levels satisfy*

$$e^{FB} \geq \bar{e}_2^C(1) > \bar{e}_1^C > \bar{e}_2^C(0) = e^{SB}.$$

b) If $I_2 > P(e^{SB})$ *and* $I_1 \leq I_1^T(I_2)$*, then the project is installed and it is terminated whenever the first period was a failure,* $x_1 = x_2(1) = 1$*,* $x_2(0) = 0$*, and*

$$e^{FB} \geq e_2^T(1) > e_1^T > e^{SB}.$$

c) Otherwise, the project is not installed at all, $x_1 = 0$*.*

Proof. See Section 2.6 at the end of the chapter.

As we have seen in the previous section, if the project was always continued under full commitment, the principal implemented a second-period effort level smaller than e^{SB} when the first period was a failure. The resulting smaller second-period rent acted as an indirect punishment of the wealth-constrained agent for the first-period failure. This is no longer possible if renegotiation cannot be ruled out, because at date 2 the principal would prefer to implement e^{SB} in order to maximize her second-period profit. While thus the “stick” is no longer available, the principal can still make use of the “carrot;” i.e., she can indirectly reward first-period effort by implementing an effort level larger than e^{SB} following a first-period success.[18] As a result, it is still cheaper for the principal to motivate the agent to exert first-period effort in the two-period model than in the one-shot benchmark model, so that $\bar{e}_1^C > e^{SB}$.

Just as in the full commitment regime, for sufficiently large investment costs I_2, the principal would be better off if she terminated the project whenever the first-period was a failure. This is clearly the case if continuation would

[17] Note that, in particular, this means that it is inconsequential how the renegotiation surplus would be split at date 2. The principal can achieve the same outcome that would be attained if she had all bargaining power in the renegotiation game by designing the appropriate renegotiation-proof contract at the outset.

[18] Note that the principal would like to reduce her promised payment t_{11} after a first-period success has occurred (in order to implement e^{SB} in the second period), but in this case there is no scope for mutually beneficial renegotiation. The agent would insist on the original contract, which gives him a larger rent.

cause more costs than gains, $I_2 > P(\bar{e}_2^C(0)) = P(e^{SB})$, but it is also the case for smaller investment costs I_2, because the threat to terminate improves the first-period incentives. Hence, there is a cut-off level $\bar{I}_2^* < I_2^* < P(e^{SB})$, in analogy to the commitment scenario. However, if renegotiation cannot be ruled out, at date 2 the principal prefers to continue the project as long as she can make a positive second-period profit by doing so. Her threat to terminate the project after a first-period failure is no longer credible, unless her expected second-period profit in case of continuation would actually be negative, $P(e^{SB}) - I_2 < 0$.

In other words, if $\bar{I}_2^* < I_2 < P(e^{SB})$, the principal would like to commit to termination following a first-period failure, but she cannot do so. This observation has peculiar implications with regard to the project that the principal will choose at the outset, as is highlighted in Corollary 3 below. A new kind of inefficiency occurs, which we saw neither in the well-known one-shot problem nor in the two-period model with full commitment.

Corollary 2.4. *Assume that the principal cannot commit not to renegotiate and $I_1 < \bar{I}_1^C(P(e^{SB}))$. If, ceteris paribus, I_2 is increased, then the principal's expected profit is reduced, except for the point $I_2 = P(e^{SB})$. At this point, there is an upward jump, which is bounded from below by $e^{SB}A(e^{SB})$.*

Proof. See Section 2.6 at the end of the chapter.

Corollary 2 says that the principal can be better off if her continuation costs I_2 are increased, which may be surprising at first sight. Yet, this result follows immediately from the fact that the principal would like to commit to termination after a first-period failure for all $I_2 \geq \bar{I}_2^*$, but given that renegotiation cannot be ruled out, she can do so only when $I_2 \geq P(e^{SB})$. Hence, her expected profit makes an upward jump at $I_2 = P(e^{SB})$. This effect can be so strong that the principal would even prefer to have higher investment costs in both periods, or similarly, she would prefer to install a project that can only yield a smaller revenue R.

Corollary 2.5. *Assume that the principal cannot commit not to renegotiate. If at date 0 the principal can choose between two (otherwise identical) projects with possible returns $R = R_g$ and $R = R_b < R_g$, she may prefer the bad project that can yield R_b only.*

Proof. See Section 2.6 at the end of the chapter.

For example, let $c(e) = \frac{1}{2}e^2$, $I_2 = 0.12$, $R_b = 0.68$, and $R_g = 0.7$. It is straightforward to show that the principal's expected profit is $\Pi \approx 0.147 - I_1$

if she installs the "good" project that can yield R_g, while it is $\Pi \approx 0.157 - I_1$ if she installs the "bad" project that can yield R_b only (and is otherwise identical). Note that if $I_1 = 0.15$, this even means that while the principal would be willing to install the "bad" project, the "good" project would never be funded.

Intuitively, pursuing a bad project that can yield a relatively small return (or, similarly, hiring a less qualified agent who can generate only a small return or who requires higher investments by the principal) acts as a commitment device. The principal knows that if she chooses the more attractive alternative, then at date 2 she cannot resist the temptation to continue after a first-period failure. For this reason, a project can be just "too good" to be funded or an "overqualified" agent may not be hired.[19]

2.5 Conclusion

In this paper, we have extended the literature on repeated moral hazard problems to cover hidden action models in which the agent is risk-neutral but wealth-constrained. We have compared the induced effort levels across periods and states. Moreover, we have identified a novel kind of potential inefficiency that has escaped the previous literature.

The present contribution seems to be sufficiently simple to be used as a building block in more applied work. As has been pointed out in the introduction, our model shares some features with dynamic adverse selection models. It might thus be applied in fields which previously have been studied from the perspective of the literature on precontractual private information and soft budget constraints. Specifically, applications of our model could help to explain the funding of inferior projects (e.g., in the context of development aid), even if the project quality is commonly known. Our model could also be applied in the field of corporate finance, where moral hazard problems with risk-neutral but wealth-constraint agents are ubiquitous (see Tirole (2005)).[20]

[19] Lewis and Sappington (1993) have also pointed out that employers will sometimes not hire applicants who are "overqualified," even when their salary expectations are modest. However, their model is quite different from ours; they consider an adverse selection problem with countervailing incentives due to type-dependent reservation utilities. Note that in our model a more productive agent might not be hired even if his reservation utility is not higher than the one of a less qualified agent.

[20] On a less serious note, the model could be applied to dating and marriages. If one's dream partner knows that anything will be forgiven, he or she may spent less effort to remain faithful. It may therefore not be a good idea to "date above one's level."

It is straightforward to relax several assumptions that were made to keep the exposition as clear as possible. For example, if it is required by an application, one might easily generalize the model by allowing different cost functions and different returns in the two periods. Moreover, one can dispense with the assumption that the principal has all bargaining power. Regardless of the bargaining protocol, the principal would only be willing to participate if her investment costs were covered. Hence, qualitatively our main findings would still be relevant. One could also consider the case in which the agent's wealth or his reservation utility may be positive. As long as his reservation utility is smaller than his rent, nothing changes. As long as the agent is not wealthy enough to "buy the firm," the effects highlighted in our model continue to be relevant.

2.6 Proofs

Proof of Proposition 1.
The proof proceeds in several steps.

Step 1. Consider first the case $x_1 = x_2(0) = x_2(1) = 1$, so that the project is always continued. The incentive compatibility constraints can then be written as

$$\begin{aligned} c'(e_2(0)) &= t_{01}, \\ c'(e_2(1)) &= t_{11}, \\ c'(e_1) &= t_1 + A(e_2(1)) - A(e_2(0)). \end{aligned}$$

The principal chooses the payments or, equivalently, the effort levels to maximize her expected payoff

$$e_1 \left[R - t_1 + P(e_2(1))\right] + (1 - e_1)P(e_2(0)) - I_2 - I_1,$$

where $t_1 = c'(e_1) - A(e_2(1)) + A(e_2(0))$ must be nonnegative. Thus, Π is a continuous function defined on the compact set $[0,1]^3 \cap \{t_1 \geq 0\}$ and as such it must have a maximum. The conditions on c ensure that the maximum is not on the boundary of $[0,1]^3$, but it could be on the boundary of $\{t_1 \geq 0\}$. The Kuhn-Tucker necessary condition for a solution of the maximization problem

$$\max_{e_1, e_2(0), e_2(1)} P(e_1) + e_1 \left[S(e_2(1)) - S(e_2(0))\right] + P(e_2(0)) - I_2 - I_1$$

subject to $t_1 = c'(e_1) - A(e_2(1)) + A(e_2(0)) \geq 0$ is

$$\nabla \Pi + \lambda \nabla t_1 = 0 \quad \text{with } \lambda \geq 0 \text{ and } \lambda > 0 \Rightarrow t_1 = 0.$$

This gives us the three equations

$$P'(e_1^C) + S(e_2^C(1)) - S(e_2^C(0)) = -\lambda c''(e_1^C), \tag{2.1}$$

$$e_1^C S'(e_2^C(1)) = \lambda A'(e_2^C(1)), \tag{2.2}$$

$$P'(e_2^C(0)) - e_1^C S'(e_2^C(0)) = -\lambda A'(e_2^C(0)). \tag{2.3}$$

Furthermore, if $\lambda > 0$, then $t_1 = 0$ and using equation (2.1), we can calculate that

$$\lambda = e_1^C - \frac{R + P(e_2^C(1)) - P(e_2^C(0))}{c''(e_1^C)}.$$

Since P is bounded by $S(e^{FB}) < R$, this means that even if λ is positive, $\lambda < e_1^C$.

From equation (2.2), we see that

$$e_2^C(1) \le e^{FB} \text{ and } e_2^C(1) = e^{FB} \Leftrightarrow \lambda = 0.$$

This equation can also be written as

$$P'(e_2^C(1)) = \frac{\lambda - e_1^C}{e_1^C} A'(e_2^C(1)) < 0.$$

Thus, we can infer $e_2^C(1) > e^{SB}$. Likewise, equation (2.3) can be rearranged to

$$P'(e_2^C(0)) = \frac{e_1^C - \lambda}{1 - e_1^C} A'(e_2^C(0)) > 0.$$

Therefore, $e_2^C(0) < e^{SB}$. Moreover, equation (2.1) implies $e_1^C > e^{SB}$, because

$$P'(e_1^C) = S(e_2^C(0)) - S(e_2^C(1)) - \lambda c''(e_1^C) < 0.$$

Finally, $e_1^C < e_2^C(1)$ follows from

$$c'(e_1^C) = A(e_2^C(1)) - A(e_2^C(0)) < c'(e_2^C(1))$$

in the case $\lambda > 0$ and from

$$R - c'(e_1^C) = e_1^C c''(e_1^C) - S(e^{FB}) + S(e_2^C(0)) > t_1 + P(e_2^C(0)) > 0$$

(using the fact that $ec''(e) > c'(e)$ for $e > 0$) in the case $\lambda = 0$. Thus, we have proved that $e^{FB} \ge e_2^C(1) > e_1^C > e^{SB} > e_2^C(0) > 0$. The principal's expected profit in the case under consideration is

$$\Pi^C(I_1, I_2, R) = P(e_1^C) + e_1^C \left[S(e_2^C(1)) - S(e_2^C(0))\right] + P(e_2^C(0)) - I_2 - I_1.$$

Step 2. Consider next the case in which the principal chooses $x_1 = x_2(1) = 1$, $x_2(0) = 0$, so that the project is terminated whenever the first period was a failure. The optimal contract can then be characterized in analogy to Step 1. In particular, the first-order conditions now read

$$P'(e_1^T) + S(e_2^T(1)) - I_2 = -\lambda c''(e_1^T),$$
$$e_1^T S'(e_2^T(1)) = \lambda A'(e_2^T(1)),$$

and $t_1 = c'(e_1^T) - A(e_2^T(1)) \ge 0$ is binding if $\lambda > 0$. In analogy to Step 1, it follows that $e^{FB} \ge e_2^T(1) > e_1^T > e^{SB}$.[21] The expected profit in this case is

$$\Pi^T(I_1, I_2, R) = P(e_1^T) + e_1^T \left[S(e_2^T(1)) - I_2\right] - I_1.$$

[21] Note that to show $e_1^T > e^{SB}$, we now use $S(e_2^T(1)) > I_2$, which due to $I_2 \le I_1$ must hold if the case under consideration is more profitable for the principal than $x_1 = 0$.

Step 3. It is straightforward to check that $x_1(1) = 0$, $x_2(0) = 1$ can never be optimal. Among the remaining alternatives (the cases discussed in Steps 1 and 2 and $x_1 = 0$), the principal will always commit to the one leading to the largest expected profits. For a given R, the expected profits $\Pi^C(I_1, I_2, R)$ and $\Pi^T(I_1, I_2, R)$ are decreasing in I_1 and I_2. In particular, using the envelope theorem it follows that

$$\frac{d\Pi^T(I_1, I_2, R)}{dI_2} = -e_1^T > -1 = \frac{d\Pi^C(I_1, I_2, R)}{dI_2}.$$

Note that $\Pi^C(I_1, 0, R) - \Pi^T(I_1, 0, R)$ does not depend on I_1 and is strictly positive, because in Step 1 we showed that it is not optimal to set $e_2^C(0) = 0$. Moreover, if I_2 were prohibitively large, continuation would never be optimal. Hence, there exists a unique cut-off level $I_2^* > 0$, such that $\Pi^C(I_1, I_2^*, R) = \Pi^T(I_1, I_2^*, R)$. If $I_2 \leq I_2^*$, continuation is better than termination, and if in addition $I_1 \leq I_1^C(I_2) := \Pi^C(0, I_2, R)$, then $x_1 = 1$ is optimal. If $I_2 > I_2^*$ and $I_1 \leq I_1^T(I_2) := \Pi^T(0, I_2, R)$, then the project is started and it is terminated following a first-period failure. Otherwise, the project is not installed. It remains to be shown that the region in which the termination contract is optimal is nonempty, given that $I_2 \leq I_1$. As has already been pointed out in the discussion following Proposition 1, $I_2^* < P(e_2^C(0))$.[22] Hence,

$$I_1^C(I_2^*) = \Pi^C(0, 0, R) - I_2^* > 2P(e^{SB}) - P(e^{SB}) > I_2^*,$$

and the proposition has been proved.

Proof of Corollary 1.

Consider the optimal contract in the case of the bad project ($R = R_b$). In the case of the good project ($R = R_g$), the principal could simply offer the same contract. Then the agent's behavior would be the same, but the principal's expected profit would be larger. By optimally adjusting the contract in the case of the good project, the principal can do even better.

Proof of Proposition 2.

The proof proceeds again in several steps.

Step 1. The principal now has to take into consideration additional renegotiation-proofness constraints. First, if the project is continued, then $e_2(y_1) \in [e^{SB}, e^{FB}]$ must be satisfied. If the original contract induced $e_2(y_1) < e^{SB}$, then at date 2 both parties could be made better off if the transfer payments were renegotiated such that the agent chooses e^{SB}. Similarly, implementing $e_2(y_1) = e^{FB}$ would make both parties better off if the original contract induced $e_2(y_1) > e^{FB}$. If $e_2(y_1) \in [e^{SB}, e^{FB}]$, no renegotiation occurs,

[22] To see this formally, note that $\Pi^C(I_1, P(e_2^C(0)), R) < P(e_1^C) + e_1^C \left[S(e_2^C(1)) - P(e_2^C(0))\right] - I_1 < \Pi^T(I_1, P(e_2^C(0)), R)$.

because effort levels closer to e^{FB} are more efficient, but given that transfer payments to the agent must be non-negative, the principal's second-period profit is larger the closer the induced effort level is to e^{SB} (see Section 2.1). Second, if the project is installed and $I_2 \leq P(e^{SB})$, then $x_2(y_1) = 1$ must be satisfied (if $x_2(y_1) = 0$, renegotiation would occur, because at date 2 both parties could be made better off by continuing the project). If $I_2 > P(e^{SB})$, then $x_2(y_1) = 0$ is renegotiation-proof, whereas $x_2(y_1) = 1$ is renegotiation-proof whenever $I_2 \leq S(e_2(y_1))$.

Step 2. Consider the case $I_2 \leq P(e^{SB})$. Given that the project is installed, the analysis is analogous to Step 1 of the proof of Proposition 1, where $x_1 = x_2(0) = x_2(1) = 1$. In particular, equation (2.3) now has to be replaced with $\bar{e}_2^C(0) = e^{SB}$, since the renegotiation-proofness constraint $\bar{e}_2^C(0) \geq e^{SB}$ is always binding (among all renegotiation-proof effort levels, e^{SB} not only maximizes the principal's second-period expected profit, but it also provides highest incentives for the agent in the first period). The other two effort levels, $\bar{e}_1^C$ and $\bar{e}_2^C(1)$, have to be adjusted accordingly. They can be derived from the first order conditions (2.1) and (2.2), and the implications for the ordering of the effort levels remain true. The principal's expected profit in the case of unconditional continuation is

$$\bar{\Pi}^C(I_1, I_2, R) = P(\bar{e}_1^C) + \bar{e}_1^C \left[S(\bar{e}_2^C(1)) - S(e^{SB})\right] + P(e^{SB}) - I_2 - I_1.$$

The project will not be installed ($x_1 = 0$) if $I_1 > \bar{I}_1^C(I_2) := \bar{\Pi}^C(0, I_2, R)$.

Step 3. Consider next the case $I_2 > P(e^{SB})$. Since then $I_2 > I_2^*$, the termination contract ($x_1 = x_2(1) = 1$, $x_2(0) = 0$) characterized in Step 2 of Proposition 1 solves the principal's maximization problem, given that $x_1 = 1$. This contract is renegotiation-proof, because $e_2^T(1) > e^{SB}$ and $I_2 \leq I_1 \leq P(e_1^T) + e_1^T \left[S(e_2^T(1)) - I_2\right] \leq S(e_2^T(1))$, given that the principal decides to install the project. The project will not be installed if $I_1 > I_1^T(I_2)$. Finally, note that

$$I_1^T(P(e^{SB})) > \bar{I}_1^C(P(e^{SB})) = \bar{\Pi}^C(0, 0, R) - P(e^{SB}) > P(e^{SB}),$$

so that the proposition follows immediately.

Proof of Corollary 2.

$\bar{\Pi}^C(I_1, I_2, R)$ and $\Pi^T(I_1, I_2, R)$ are continuous and decreasing in I_2. At $I_2 = P(e^{SB}) > I_2^*$, we know that $\Pi^T(I_1, I_2, R) > \Pi^C(I_1, I_2, R) > \bar{\Pi}^C(I_1, I_2, R)$. Hence, given that the project is installed, at $I_2 = P(e^{SB})$ the principal's expected profit as characterized in Proposition 2 is discontinuous, and the size of the jump is given by

$$\begin{aligned}
&\Pi^T(I_1, P(e^{SB}), R) - \bar{\Pi}^C(I_1, P(e^{SB}), R) \\
&= P(e_1^T) + e_1^T \left[S(e_2^T(1)) - P(e^{SB})\right] - \left(P(\bar{e}_1^C) + \bar{e}_1^C \left[S(\bar{e}_2^C(1)) - S(e^{SB})\right]\right) \\
&> P(\bar{e}_1^C) + \bar{e}_1^C \left[S(\bar{e}_2^C(1)) - P(e^{SB})\right] - \left(P(\bar{e}_1^C) + \bar{e}_1^C \left[S(\bar{e}_2^C(1)) - S(e^{SB})\right]\right) \\
&> e^{SB}A(e^{SB}).
\end{aligned}$$

Proof of Corollary 3.
Take any $R = \tilde{R}$ and I_1, I_2 such that $I_2 = P(e^{SB})$ and $I_1 \leq I_1^T(I_2)$. Corollary 2 shows that $\Pi^T(I_1, I_2, \tilde{R}) > \bar{\Pi}^C(I_1, I_2, \tilde{R}) + e^{SB}A(e^{SB})$. Note that, using the envelope theorem, $dP(e^{SB})/dR = e^{SB} > 0$. Proposition 2 thus implies that if R_g is slightly larger than $\tilde{R}$ and R_b is slightly smaller than $\tilde{R}$, then the principal prefers R_b (leading to the expected profit $\Pi^T(I_1, I_2, R_b)$) to R_g (leading to the expected profit $\bar{\Pi}^C(I_1, I_2, R_g)$).

3

Expectation damages, divisible contracts, and bilateral investment

3.1 Introduction

Real-world contracts sometimes look surprisingly simple given the complexity of the environment. A production contract between a buyer and a seller might specify only a fixed quantity and a price per unit, although between the signing of the contract and actual trade many contingencies may arise that affect the value of trade. As a consequence, after the uncertainty about valuation and production costs is resolved, the parties might observe that they can make a larger profit by trading more or less than the stipulated quantity. In this case, they can resort to renegotiation to reach a trade decision that is optimal given the relationship-investments they made. The main purpose of their contract is then to set the right incentives to invest.

The contract affects the investment incentives by serving as a disagreement point in the renegotiations. Here, standard breach remedies play a role, since the payoff that one party can realize unilaterally also depends on the consequences of breach. In this chapter, we focus on the standard expectation damages remedy. It is assumed that courts can verify all relevant information to award expectation damages, which compensate the victim of breach for the loss of profit. Our result is that contracts do not need to be contingent under this damage rule: price and quantity can be adjusted to induce both parties to invest efficiently.

Two effects of renegotiation and standard breach remedies on investment have been identified in the literature. If the parties leave the trade decision to ex post negotiations, they will underinvest relative to the first best, due to a hold-up problem (see e.g. Williamson (1985), Hart and Moore (1988)). In contrast, the economic analysis of breach remedies by Shavell (1980,1984), Rogerson (1984) and others reveals that if contracts and breach remedies are available then there can also be an overinvestment effect. Remedies such as the expectation measure act as an insurance against breach. The victim of breach invests more than if he or she internalized the lost investment in case of breach.

These two intuitions are integrated by Edlin and Reichelstein (1996) (henceforth ER), who show that it is possible to balance the hold-up effect against the overinvestment effect when contracts are enforced by the standard breach remedies of expectation damages or specific performance. They find that a continuous quantity in the contract is a powerful tool to adjust the investment incentives of one party.[1] In addition, when the breach remedy is specific performance, the incentives of two parties can be aligned with a single quantity if the payoff functions have a particular form. In contrast, for the regime of expectation damages ER show that for a deterministic and linear cost function no fixed price-quantity contract exists that achieves first best investment decisions. They conclude that specific performance is better suited for two-sided investment problems than expectation damages.

The present chapter extends the analysis to more general cost and valuation functions and finds that the scope of expectation damages to solve a bilateral hold-up problem is much larger than this counterexample suggests. It turns out that in ER's framework with divisible contracts[2], the per-unit price can be used as an additional instrument to fine-tune both parties' incentives to invest. ER establish their positive results for one-sided investment through the adjustment of quantity alone, while the price is set high or low such that always the same party breaches. With an intermediate price, any party may breach the contract, and the parties' probabilities of breach vary with price. Therefore, a contract that specifies an up-front transfer, a quantity, and a per-unit price often suffices to obtain efficient two-sided investment.

The reason why the first best can not be obtained under the expectation damages remedy in the case of constant and deterministic marginal cost is that the seller's investment decision completely determines who breaches. The breaching party never gains from having invested more than the efficient amount, hence the hold-up effect dominates this party's incentives. In order to balance both investment decisions, both parties must face the risk of breach.

[1] The idea that a continuous variable can help to reach the first best by balancing one party's investment incentives also appears in related articles such as Chung (1991), Aghion, Dewatripont and Rey (1994), and Nöldeke and Schmidt (1995). In these papers, the first best can be reached because renegotiation leaves one party with the full surplus.

[2] A divisible contract consists of several items and the price to be paid is apportioned to each item. It can be broken into its component parts, such that each unit together with the per-unit price can be treated as a separate contract, which is fulfilled or breached independently.

While no single per-unit price has this effect in ER's example, we demonstrate that the parties can use a lottery between an extremely high and extremely low per-unit price. Like the intermediate per-unit price, the probability of a high price can be used as an additional continuous variable to fine-tune investment incentives.

The chapter is organized as follows: Section 3.2 introduces the model, while in Section 3.3 the ex post consequences of expectation damages with divisible contracts are discussed. The main result on the optimality of price-quantity contracts is presented in Section 3.4, and Section 3.5 deals with stochastic prices. Concluding remarks can be found in Section 3.6. Proofs have been relegated to the end of the chapter.

3.2 The model

The sequence of events is illustrated in Figure 3.1. A seller and a buyer, both of whom are risk-neutral, have to incur relationship-specific investments in preparation of future trade. To protect these investments, they sign a contract at date 1, specifying a per-unit price $\bar{p}$ and a quantity $\bar{q} \in [0, q^{max}]$ of the good to be traded.[3] The parties may also exchange an up-front transfer T to divide the gains from trade after price and quantity are chosen to maximize joint surplus.

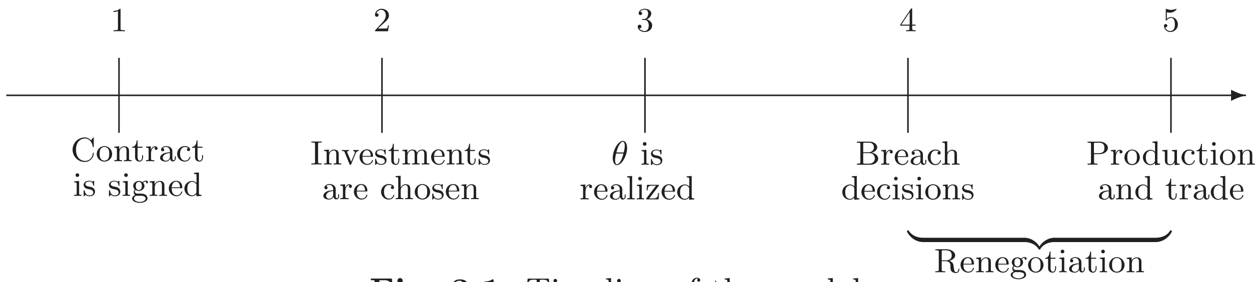

Fig. 3.1. Timeline of the model.

At date 2, the seller invests to decrease his marginal cost, and the buyer invests to increase her marginal benefit of the good. The costs of their investments (or reliance expenditures, in legal terms) are denoted by $\sigma \in [0, \sigma^{max}]$

[3] The quantity of the traded good could also be interpreted as duration of the business relationship. With this interpretation the problem has a dynamic structure which is analyzed by Guriev and Kvasov (2005).

and $\beta \in [0, \beta^{max}]$, respectively. These investment decisions may be difficult to describe or to observe, and consequently are not contractible. The exact shape of the cost and valuation functions becomes commonly known at date 3, when the state of the world $\theta \in \Theta$ is realized. The contingency θ reflects exogenous uncertainty and is drawn from a compact state space $\Theta \subset \mathbb{R}^n$ according to a distribution function F.

At date 4, both seller and buyer decide whether they want to breach to a quantity lower than the one specified in the contract. The consequences of breach are determined by expectation damages, either as the default breach remedy or because the contract explicitly specifies this breach remedy. A more detailed description of the consequences of breach follows in Section 3.3. The payoff as determined by the legal consequences constitutes the disagreement point in subsequent renegotiations. The outcome of the negotiations is assumed to be the (generalized) Nash bargaining solution, where $\gamma \in [0, 1]$ denotes the seller's bargaining power. We use the following notation and assumptions:

- The seller's payoff of producing quantity q is $-C(\sigma, \theta, q) - \sigma$. The cost function C is increasing and strictly convex in q, and $(\sigma, q) \mapsto C(\sigma, \theta, q)$ is twice continuously differentiable for all $\theta \in \Theta$, with $C_{\sigma q} \leq 0$. The functions C and C_σ are assumed to be continuous in θ.
- The buyer's payoff of obtaining quantity q is $V(\beta, \theta, q) - \beta$. The valuation function V is increasing and strictly concave in q, and $(\beta, q) \mapsto V(\beta, \theta, q)$ is twice continuously differentiable for all $\theta \in \Theta$, with $V_{\beta q} \geq 0$. Moreover, V and V_β are continuous in θ.

Ex post, trade of a quantity q creates a joint surplus of

$$W(\beta, \sigma, \theta, q) := V(\beta, \theta, q) - C(\sigma, \theta, q).$$

This is maximized at the efficient quantity

$$Q^*(\beta, \sigma, \theta) := \underset{q \in [0, q^{max}]}{\arg\max}\, W(\beta, \sigma, \theta, q).$$

We denote by (β^*, σ^*) the efficient investment levels which maximize

$$\int W(\beta, \sigma, \theta, Q^*(\beta, \sigma, \theta)) dF - \beta - \sigma.$$

in $[0, \beta^{max}] \times [0, \sigma^{max}]$. We assume that these are the unique maximizers.

3.3 Breach decisions

The contract and the legal consequences define a game between seller and buyer, which we will solve by backward induction. In this section we analyze the ex post subgame, when cost and valuation functions are realized and observable by both players and the court. Renegotiation will lead the parties to the ex post efficient trade decision. In order to determine their payoffs, we first explore the consequences of breach to see how much a party can achieve unilaterally. We abstract from issues like litigation costs or difficulties to assess damages.

According to the expectation damage rule, the breaching party has to compensate the victim of breach for the loss caused by the breach. The goal of this remedy is to put the victim in as good a position as if the contract had been fulfilled. This rule is, however, not applied literally in cases where this party faces negative profits from completion of the contract. If one party's breach is advantageous for the other, it is not possible to sue for a reward, but damage payments are zero.

The buyer and the seller face symmetric decisions in this subgame. While the seller can breach by producing and delivering less than ordered, the buyer can breach by announcing her breach decision before the unwanted units are produced. In that case, the seller can only recover the profit margin of the canceled goods, but not their cost of production. Suppose that first the buyer announces her anticipatory breach decision $q_B \leq \bar{q}$, followed by the seller's announcement $q_S \leq \bar{q}$.[4] Then the seller subsequently delivers the quantity $q = \min\{q_S, q_B\}$, as he will not be compensated for producing a larger quantity.

Since the contract explicitly specifies a price per unit, it is divisible, and the seller is entitled to a payment of $\bar{p}q$ for his partial performance. The payoffs from this part of the contract are

$$S(\sigma, \theta, q) := \bar{p}q - C(\sigma, \theta, q)$$

and

$$B(\beta, \theta, q) := V(\beta, \theta, q) - \bar{p}q.$$

Damages are confined to that part of the contract that was breached. If $q < q_B$, the seller is liable for the buyer's loss on the units between q and q_B. On all units above q_B the contract counts as consensually canceled. The seller has to pay

$$\max\{B(\beta, \theta, q_B) - B(\beta, \theta, q), 0\}^5$$

in damages to the buyer. Similarly, if $q = q_B$, the seller can sue the buyer for the sum

[4] The order of announcements does not matter for the outcome of the subgame.

[5] This is the formula used by ER and most of the literature. Another approach is to compare the utility resulting from breach with a hypothetical, "reasonably foreseeable" (Cooter (1985)) valuation $B(\beta^*, \theta, q)$ instead of the actual valuation. As shown by Leitzel (1989) and Schweizer (2005) this solves the overreliance problem.

$$\max\{S(\sigma,\theta,q_S) - S(\sigma,\theta,q),0\}.$$

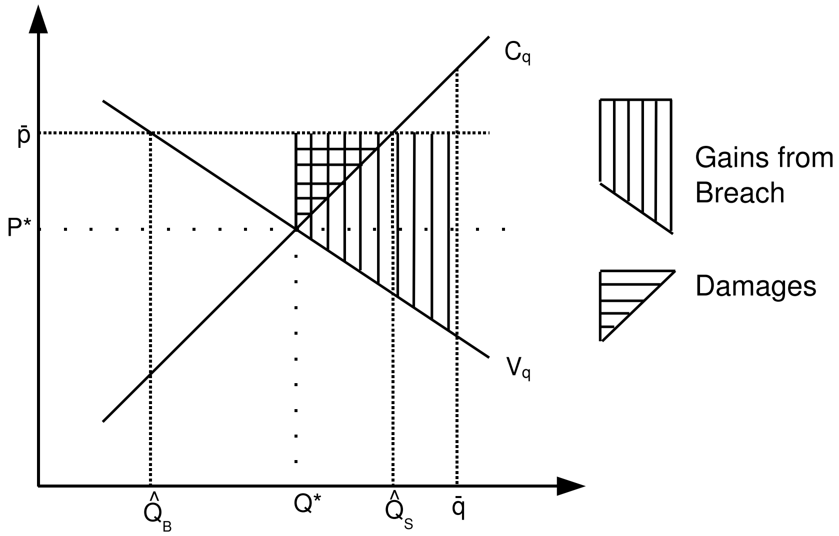

Fig. 3.2. This figure shows that at Q^*, the buyer's marginal damage payments equal her marginal gain from breach.

We will show that with these damage payments the *efficient breach* property of expectation damages continues to hold.[6] An intuition for this result can be gained from Figure 3.2 which shows marginal cost and valuation. The two functions intersect at the efficient quantity Q^*, and in the case that is illustrated, this quantity is lower than the contracted quantity $\bar{q}$, and the "equilibrium price" P^* is lower than the contracted price $\bar{p}$. No damage is done by breach on the units above $\hat{Q}_S$, and no damages have to be paid. For breach on all other units, the buyer has to pay the difference between $\bar{p}$ and the "supply curve" to the seller, while she gains the difference between $\bar{p}$ and the "demand curve". The buyer thus breaches on all units in excess of the quantity Q^*, where these differences become equal.

To formalize this result we define two quantities that are related to supply and demand at the contract price:

$$\hat{Q}_S := \arg\max_{q\leq\bar{q}} S(\sigma,\theta,q) = \min\{\bar{q}, \arg\max_{q} S(\sigma,\theta,q)\}$$

and

[6] This would not be true if the court ignored the higher breach quantity and calculated damages with respect to the contracted quantity. Such a rule would aggregate gains and losses on breached units, such that in some contingencies breach leads to a less than efficient quantity.

$$\hat{Q}_B := \arg\max_{q \leq \bar{q}} B(\beta, \theta, q) = \min\{\bar{q}, \arg\max_{q} B(\beta, \theta, q)\},$$

where the second equality is due to concavity of the functions S and B. We also define

$$P^* := C_q(\sigma, \theta, Q^*).$$

Let us assume first that $Q^* \leq \hat{Q}_S$, which is equivalent to the case that $Q^* \leq \bar{q}$ and $P^* \leq \bar{p}$. We will show that equilibrium payoffs are equal to $S(\sigma, \theta, \hat{Q}_S)$ for the seller and $W(\beta, \sigma, \theta, Q^*) - S(\sigma, \theta, \hat{Q}_S)$ for the buyer. If the seller chooses $q_S = \hat{Q}_S$, the damage rule ensures him a payoff of $S(\sigma, \theta, \hat{Q}_S)$ plus a possible gain from renegotiation. This is true regardless of the buyer's breach decision, since even if the seller turned out to be the one to breach, he would pay no damages. On the other hand, if the buyer chooses the anticipatory breach decision $q_B = Q^*$, she gets at least $W(\beta, \sigma, \theta, Q^*) - S(\sigma, \theta, \hat{Q}_S)$. To see that, note that she gets $W(\beta, \sigma, \theta, Q^*) - S(\sigma, \theta, q_S)$ if she pays positive damages and at least $B(\beta, \theta, Q^*)$ if she has to pay no damages.

Therefore, in any Nash equilibrium the buyer gets at least $W(\beta, \sigma, \theta, Q^*) - S(\sigma, \theta, \hat{Q}_S)$, while the seller gets at least $S(\sigma, \theta, \hat{Q}_S)$. Since in sum the payoffs cannot be higher than the maximal joint surplus, these must be the parties' equilibrium payoffs, achieved by $q_B = Q^*$ and $q_S = \hat{Q}_S$. In the appendix, we show that these strategies form the unique equilibrium if $\gamma \in (0, 1)$. The key intuition here is that the breaching party receives all gains from the breach decision.

If $Q^* \leq \hat{Q}_B$, which is equivalent to the case that $Q^* \leq \bar{q}$ and $P^* \geq \bar{p}$, the buyer makes a positive profit on every unit that is efficient to trade. This time the seller breaches to $q_S = Q^*$ and pays damages to the buyer. Their payoffs are $B(\beta, \theta, \hat{Q}_B)$ for the buyer and $W(\beta, \sigma, \theta, Q^*) - B(\beta, \theta, \hat{Q}_B)$ for the seller.

In the case $Q^* > \bar{q}$ the optimal quantity can be reached by renegotiation only. In these contingencies, one of the parties makes a profit on each unit traded under the contract. The analysis here is the same as in ER. In this case, there will be no breach, and renegotiation can take place before or after delivery of $\bar{q}$. The parties share the additional gain from efficient trade

$$\Delta(\beta, \sigma, \theta, \bar{q}) := W(\beta, \sigma, \theta, Q^*) - W(\beta, \sigma, \theta, \bar{q})$$

according to bargaining power, leaving the seller with a share of γ, the buyer with a share of $1 - \gamma$ of the renegotiation surplus. Hence, their payoffs are $S(\sigma, \theta, \bar{q}) + \gamma \Delta(\beta, \sigma, \theta, \bar{q})$ for the seller and $B(\beta, \theta, \bar{q}) + (1 - \gamma)\Delta(\beta, \sigma, \theta, \bar{q})$ for the buyer.

3.4 Optimal contracts

In this section, we turn to the ex ante perspective of the game and analyze the investment choices that are induced by the contract. Taking into account the possible ex post equilibrium payoffs as described in the last section we obtain the following expression for the seller's expected payoff:

$$\begin{aligned} s(\beta, \sigma) = & \int_{[Q^* > \bar{q}]} S(\sigma, \theta, \bar{q}) + \gamma \Delta(\beta, \sigma, \theta, \bar{q}) dF \\ & + \int_{[\hat{Q}_B \geq Q^*]} W(\beta, \sigma, \theta, Q^*) - B(\beta, \theta, \hat{Q}_B) dF \\ & + \int_{[\hat{Q}_S \geq Q^*]} S(\sigma, \theta, \hat{Q}_S) dF - \sigma. \end{aligned} \tag{3.1}$$

The buyer's expected payoff, denoted by $b(\beta, \sigma)$, is derived analogously to the seller's expected payoff. The payoff functions are easiest to analyze for extreme contracts, for which at efficient investment at least one of the events "renegotiation", "buyer breaches" and "seller breaches" never occurs. We define

$$q_H := \max_\theta Q^*(\beta^*, \sigma^*, \theta), \; q_L := \min_\theta Q^*(\beta^*, \sigma^*, \theta)$$

and

$$p_L := \min_{\theta, \beta, \sigma} P^*(\beta, \sigma, \theta), \; p_H := \max_{\theta, \beta, \sigma} P^*(\beta, \sigma, \theta).$$

Moreover, let

$$\sigma_S(q, p) := \arg\max_\sigma s(\beta^*, \sigma)$$

denote the seller's best response to β^* and

$$\beta_B(q, p) := \arg\max_\beta b(\beta, \sigma^*)$$

the buyer's best response to σ^* if the contract specifies a quantity q and a price p.

Lemma 3.1. *For all prices p, it holds that*

$$\max \sigma_S(q_L, p) \leq \sigma^* \leq \min \sigma_S(q_H, p)$$

and

$$\max \beta_B(q_L, p) \leq \beta^* \leq \min \beta_B(q_H, p).$$

Moreover, $\sigma_S(q_H, p_L) = \{\sigma^*\}$ and $\beta_B(q_H, p_H) = \{\beta^*\}$.

PROOF: See Section 3.7 at the end of the chapter.

The intuition is that, given efficient investment, a contracted quantity as low as q_L means that the contract will always be renegotiated to a higher quantity. In the renegotiations, a party receives only a fraction of the surplus generated by the investment, therefore both parties underinvest (hold-up effect). A high contracted quantity q_H means renegotiation never occurs, and being sometimes the non-breaching party induces the parties to prepare for trade of a high quantity. Both parties overinvest, except for the case of a very high or very low price: In this case one party always breaches and invests efficiently, in anticipation of the efficient breach decision. This last result has been studied in detail in Edlin (1996), who uses the term "Cadillac contract" for a contract that is always breached.

This extreme kind of contracts will only in exceptional cases be able to induce efficient investment in equilibrium.[7] To infer from the behavior of best responses for extreme contracts to the behavior for contracts with an intermediate price and quantity, we need a continuity assumption. In line with the analysis of the one-sided investment case in ER, we make the following assumption.

ASSUMPTION 2 *The best response correspondences* $(q,p) \mapsto \sigma_S(q,p)$ *and* $(q,p) \mapsto \beta_B(q,p)$ *have a continuous selection.*

It would be desirable to have a characterization of the cost and valuation functions for which this condition holds. Sufficient conditions can be found, as for example the following assumption.

ASSUMPTION 3 *Let* $(\sigma, q) \mapsto C(\sigma, \theta, q)$ *be strictly convex and* $(\beta, q) \mapsto V(\beta, \theta, q)$ *be strictly concave for all* $\theta \in \Theta$.

This assumption already implies some of the assumptions that were introduced in Section 3.2, and it holds that

Lemma 3.2. *Assumption 3 implies Assumption 2.*

PROOF: See Section 3.7 at the end of the chapter.

Continuity of best responses indeed ensures existence of an optimal contract, as is illustrated in Figure 3.3, which provides the intuition for our main result:

[7] One of these special cases is $\gamma \in \{0,1\}$. It has been known since Chung (1991) that specific performance can lead to two-sided efficient investment if one party has all the bargaining power. The same is true for expectation damages if the price is set such that this party always breaches. The quantity can then be used to generate investment incentives for the other party.

Proposition 3.3. *Given Assumption 2, there exists a non-contingent contract* $(\bar{q}, \bar{p})$ *such that the first best investment levels* (β^*, σ^*) *constitute a Nash equilibrium of the induced game.*

PROOF: See Section 3.7 at the end of the chapter.

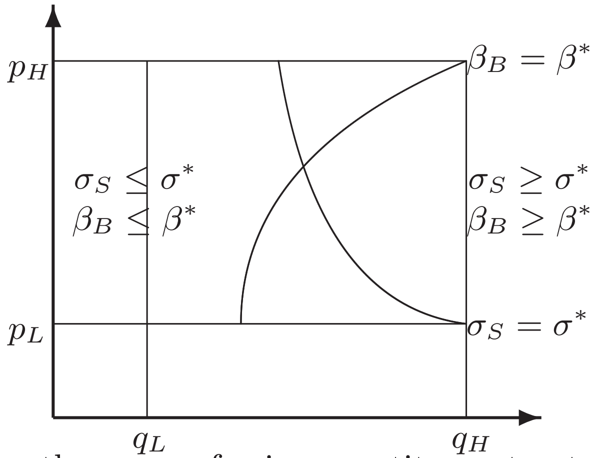

Fig. 3.3. This figure shows the space of price-quantity contracts. For low and high quantities it is indicated whether the best response to efficient investment of the other party is underinvestment, overinvestment or efficient investment. The contracts on the paths have the property that the best response to efficient investment is equal to efficient investment.

In order to find the optimal contract given a particular problem we can use the first order conditions of the parties' maximization problem. Since the derivatives of the parties' objective functions, evaluated at the efficient investments, are continuous in p and q, there is always a contract such that the first order conditions hold. Without a sufficient condition like Assumption 3, we cannot be sure that such a solution then indeed leads to a maximum and have to check the second order conditions.

The derivatives of the expected payoff functions take a particularly simple form if cost and valuation functions belong to the following class of functions:

ASSUMPTION 4

$$C(\sigma, \theta, q) = C_1(\sigma)q + C_2(\theta, q) + C_3(\sigma, \theta)$$
$$V(\beta, \theta, q) = V_1(\beta)q + V_2(\theta, q) + V_3(\beta, \theta).$$

ER show that with this functional form, the incentives for the two parties can be aligned with a single quantity if the breach remedy is specific performance.

Corollary 3.4. *If Assumption 4 holds, and* (β^*, σ^*) *is an interior solution, the first best contract* $(\bar{q}, \bar{p})$ *has to fulfill the conditions*

$$\int_{[\hat{Q}_S \geq Q^*]} (\hat{Q}_S - Q^*) dF = (1-\gamma) \int_{[Q^* > \bar{q}]} (Q^* - \bar{q}) dF \tag{3.2}$$

$$\int_{[\hat{Q}_B \geq Q^*]} (\hat{Q}_B - Q^*) dF = \gamma \int_{[Q^* > \bar{q}]} (Q^* - \bar{q}) dF, \tag{3.3}$$

where all quantities are evaluated at $\sigma = \sigma^*$ *and* $\beta = \beta^*$*.*

PROOF: See Section 3.7 at the end of the chapter.

As an example, consider the case that cost and valuation are correlated in a way that P^* is independent of θ. Balancing incentives in such a case is potentially hard to do, because the investment decisions completely determine the identity of the breaching party. Since P^* does not vary with θ, for every combination of investments one of the events $[\hat{Q}_B > Q^*]$ and $[\hat{Q}_S > Q^*]$ is empty. From the necessary conditions (3.2) and (3.3) it follows that if $\gamma \in (0,1)$, the only candidate for a first best contract is $\bar{q} = q_H$ and $\bar{p} = P^*(\beta^*, \sigma^*, \theta)$. If the payoff functions are sufficiently concave, the parties make a larger profit with efficient investment than by preparing for a higher quantity, because overinvestment leads to renegotiation in some contingencies. To make sure that this is indeed the case, we have to appeal to Assumption 3. Two explicit examples, exploring the importance of this assumption, can be found in Appendix B.

3.5 Price adjustment clauses

In this section, we explore what kind of contracts the trading partners can write if Assumption 2 and Proposition 3.3 do not hold. One of these cases is identified by ER, who show that for $\gamma \in (0,1)$ and the cost function $C(\sigma, \theta, q) = C_1(\sigma)q$ there exists no contract $(\bar{q}, \bar{p})$ that can achieve the first best if the valuation function has positive variance. With this kind of cost function the seller faces a choice between two conflicting roles: Either he invests low and breaches the contract, or he invests high and seeks damages if necessary. Whenever such a conflict leads to a discontinuity in one party's best response, the parties may have to write more complicated contracts in order to attain the first best. We will show that if the parties can stipulate a stochastic price, they can obtain first best outcomes for a larger class of payoff functions, including linear ones.

Using a lottery between a very low and a very high price instead of an intermediate price $\bar{p}$ is a more direct way to achieve breach of both parties. Let the contract condition the price on an event that occurs with probability λ independently of cost and valuation functions, such that the low price p_L is valid if the event occurs and p_H if it does not. This resembles so-called *price* *escalator clauses* or *price adjustment clauses*, which parties can use to share the risk of breach.[8]

Proposition 3.5. *Assume that the best responses are continuous in q and λ. Then there is a quantity $\bar{q} \in [q_L, q_H]$ and a $\lambda \in [0,1]$, such that a contract specifying $\bar{q}$ and a lottery over p_L with probability λ and p_H with probability $1-\lambda$ induces the first best.*

PROOF: See Section 3.7.

Since this result does not require continuity of best responses in price, it holds for a larger class of payoff functions than Proposition 3.3.[9] The optimal contract illustrates how the performance of expectation damages depends on who will breach the contract. This is especially true for the case of the payoff functions defined in Assumption 4, for which the contract takes a very intuitive form.

Proposition 3.6. *If Assumption 4 holds, a contract for $\bar{q} = \int Q^*(\beta^*, \sigma^*, \theta) dF$ at a price p_L with probability γ (the seller's bargaining power) and p_H with probability $1-\gamma$ induces the first best.*

PROOF: See Section 3.7.

As ER show, the same contracted quantity leads to efficient investment with specific performance if Assumption 4 holds. One could ask whether stochastic prices are also able to improve the performance of this breach remedy. The answer depends on the bargaining game. While ER keep the bargaining process very general in the one-sided investment analysis, for two-sided investment they also assume the Nash bargaining solution with a constant sharing rule. When the price is so high or so low that it is always in one party's interest to sue for performance, the seller's expected profit as derived in ER is:

$$\int S(\sigma, \theta, \bar{q}) + \gamma \Delta(\beta, \sigma, \theta, \bar{q}) dF - \sigma.$$

The derivative of this expression with respect to σ does not depend on the price. Hence, investment incentives can only be generated through the contracted quantity, and there is no analog to Proposition 3.5.

While the results can not be generalized as long as only quantity can be used, price always matters to some degree. How much it matters depends on

[8] Usually, one would think of price adjustment clauses as insurance against events that are correlated with either cost or valuation. Such a clause can also help to balance incentives, but the point is much simpler to make for the independent case.

[9] A sufficient condition corresponding to Assumption 3 is that $W(\beta, \sigma, \theta, q)$ is strictly concave in (σ, q) and (β, q).

the bargaining process. Here is one example to show how much the price can matter with a different bargaining solution, a two-sided offer game with outside options, as for example modeled in W. Bentley MacLeod and David M. Malcomson (1993)[10]. Applying this solution here means treating the enforcement of trade of inefficient units as an outside option. In the case that $Q^* \geq \bar{q}$ the parties would split the renegotiation surplus equally, while in the case $Q^* \leq \bar{q}$ the seller's (buyer's) outside option would always bind if the price is p_H (p_L). If the contract specifies the low price with probability λ and the high price with probability $1-\lambda$, applying an outside option bargaining solution suggests the following payoff for the seller:

$$\int_{Q^* \geq \bar{q}} S(\bar{q}) + \frac{1}{2}\Delta(\bar{q})dF + \int_{\bar{q} > Q^*} \lambda(W(Q^*) - B(\bar{q})) + (1-\lambda)S(\bar{q})dF - \sigma.$$

Since this is the same payoff as under expectation damages with Nash bargaining, Proposition 3.5 carries over to specific performance.

Option contracts

Are there other simple contracts that can reach the first best in the linear case? The deterministic case can be solved with an option contract, but in general option contracts together with expectation damages perfom poorly. This is not surprising because there is again only one instrument to adjust incentives. We define an option contract to specify an upfront payment and a per-unit price $\bar{p}$. Ex post, the buyer can order any quantity she wants at price $\bar{p}$. The seller can subsequently decide whether he wants to breach. The outcome can also be renegotiated.

This game is easy to analyze given what we already know from Section 3.3. At date 4, the buyer orders the quantity $\hat{Q}_B$ which ensures her the maximal payoff of $B(\beta, \theta, \hat{Q}_B)$ plus a possible gain from renegotiation. The seller will deliver Q^* if $Q^* \leq \hat{Q}_B$ and $\hat{Q}_B$ otherwise. The buyer will never breach, which provides an intuition for why expectation damages perform poorly with a buyer-option contract. Besides, there is only one instrument, price, to fine-tune both incentives to invest, which will only work in special cases.

Proposition 3.7. *An option contract together with expectation damages can only implement the first best if either*

(i) $\gamma = 1$, *in which case* $\bar{p}$ *is chosen such that* $\int V_\beta(\beta^*, \theta, \hat{Q}_B(\beta^*, \theta))d\pi = 0$ *at* $p = \bar{p}$, *or*

[10] This bargaining solution is also mentioned in ER as one for which the one-sided investment result still holds.

(ii) $Q^*(\sigma^*, \beta^*, \theta) = \hat{Q}_B(\sigma^*, \beta^*, \theta)$ *for almost all* θ*. With a constant per-unit price* $\bar{p}$ *and positive variance of* Q^**, this is true if and only if* $C(\sigma, \theta, q) = C_1(\sigma)q$ *and* $\bar{p} = C_1(\sigma^*)$*.*

PROOF: See Section 3.7.

3.6 Concluding remarks

We have shown that in the framework of expectation damages with bilateral investment in Edlin and Reichelstein (1996), the first best can be restored if both parties face the risk of breaching. With divisible contracts, this can always be achieved with a lottery between a high and a low per-unit price, or with a fixed price if the payoff functions are sufficiently concave. In both cases, each party's probability of breaching varies with price, such that price and quantity are sufficient to fine-tune both sides' incentives to invest. Consequently, also in this framework the trading parties can write non-contingent contracts and obtain efficient outcomes, relying on renegotiation and the standard breach remedy of expectation damages.

The advantage of an intermediate price is that depending on the move of nature, both seller and buyer will breach sometimes. Moreover, an intermediate price seems more realistic and might lead to a lower up-front payment, thus reducing the problem to design a substantial up-front transfer such that it is not touched by the breach remedy. Prerequisite for the contractual solution identified in this chapter is that the court can readily assess the damages. Even though expectation damages is a standard remedy that courts are used to deal with, there may be cases where this information is not available, and specific performance is the better choice.

Moreover, there is a general truth behind ER's inefficiency example: the expectation damage rule treats the breaching party and the party suffering from breach asymmetrically. The only contract that overcomes the hold-up problem of the breaching party specifies such a high quantity that the non-breaching party invests too much. This intuition is likely to carry over to more general settings, as long as only quantity has an effect on investment. The contribution of this chapter is to recognize that investment incentives need not be generated by quantity alone, price matters as well. It remains to be explored how much price and quantity can achieve if contracts are not divisible, or for other legal rules and contractual environments.

3.7 Proofs

Proof of uniqueness of the equilibrium in the ex post subgame.
As in the text, we consider only the case that $\hat{Q}_S > Q^*$ and assume $\gamma \in (0,1)$. We show that the equilibrium Q^* (buyer) and $\hat{Q}_S$ in fact results from iterative elimination of dominated strategies. First, for the buyer all strategies $q_B > \hat{Q}_S$ are weakly dominated by Q^*: For any such q_B and $q_S \geq q_B$ the buyer pays no damages and gets

$$B(q_B) + (1-\gamma)\Delta(q_B) = \gamma W(q_B) + (1-\gamma)W^* - S(q_B) < W(Q^*) - S(q_S).$$

For $q_S \leq Q^*$ the payoffs compare as follows

$$\max\{B(q_B), B(q_S)\} + (1-\gamma)\Delta(q_S) \leq \max\{B(Q^*), B(q_S)\} + (1-\gamma)\Delta(q_S)$$

while for the case $q_B > q_S > Q^*$ they are

$$B(q_S) + (1-\gamma)\Delta(q_S) \leq W(Q^*) - S(q_S).$$

Second, if $q_B \leq \hat{Q}_S$, the strategy $\hat{Q}_S$ is a strictly dominant strategy for the seller. To see this, take any other q_S and first the case that $q_S \geq q_B$. In this case he gets

$$\max(S(q_S), S(q_B)) + \gamma\Delta(q_B) < S(\hat{Q}_S) + \gamma\Delta(q_B).$$

Now look at the other case $q_S < q_B$. In this case, the seller gets

$$S(q_S) - \max\{B(q_B) - B(q_S), 0\} + \gamma\Delta(q_S)$$

which is equal to

$$S(q_S)(1-\gamma) + \gamma S(q_B) + \gamma\Delta(q_B) + \gamma(B(q_B) - B(q_S)) - \max\{B(q_B) - B(q_S), 0\}.$$

One can see that this payoff is smaller than

$$S(\hat{Q}_S) + \gamma\Delta(q_B).$$

Last, of course Q^* is a strictly best reply to $\hat{Q}_S$.

Proof of Lemma 3.1.
The steps of the proof are exercised in detail only for the seller's payoff function, the result for the buyer can then be derived in a similar way. In a first step, we calculate the derivative of the seller's expected profit. For this, note that as a direct application of the envelope theorem (for constrained maximization) we get for all $\theta \in \Theta$

$$\frac{\partial}{\partial \sigma} W(\beta, \sigma, \theta, Q^*(\beta, \sigma, \theta)) = -C_\sigma(\sigma, \theta, Q^*(\beta, \sigma, \theta)), \tag{3.4}$$

and

$$\frac{\partial}{\partial \sigma} S(\sigma, \theta, \hat{Q}_S(\sigma, \theta)) = -C_\sigma(\sigma, \theta, \hat{Q}_S(\sigma, \theta)). \tag{3.5}$$

Next, to calculate the derivative $\frac{\partial}{\partial \sigma} s(\beta, \sigma)$, note that for each θ the integrand in s is the piecewise defined function

$$\sigma \mapsto \begin{cases} (1-\gamma)S(\sigma, \theta, \bar{q}) + \gamma W(\beta, \sigma, \theta, Q^*) - \gamma B(\beta, \theta, \bar{q}) & \text{if } Q^* > \bar{q} \\ S(\sigma, \theta, \hat{Q}_S) & \text{if } Q^* \leq \hat{Q}_S \\ W(\beta, \sigma, \theta, Q^*) - B(\beta, \theta, \hat{Q}_B) & \text{if } Q^* \leq \hat{Q}_B \end{cases}$$

It turns out that the piecewise defined derivative of this function is continuous, i.e., the pieces of this function are joined smoothly. We assume integrability of C_σ, so that we can interchange integration and differentiation, and get:

$$\begin{aligned} \frac{\partial}{\partial \sigma} s(\beta, \sigma) &= -\int_{[Q^* > \bar{q}]} ((1-\gamma)C_\sigma(\sigma, \theta, \bar{q}) + \gamma C_\sigma(\sigma, \theta, Q^*))dF - 1 \qquad (3.6) \\ &\quad - \int_{[\hat{Q}_S \geq Q^*]} C_\sigma(\sigma, \theta, \hat{Q}_S)dF - \int_{[\hat{Q}_B \geq Q^*]} C_\sigma(\sigma, \theta, Q^*)dF \\ &= -(1-\gamma)\int_{[Q^* > \bar{q}]} \Delta_\sigma(\beta, \sigma, \theta, \bar{q})dF - \int_{[\hat{Q}_S \geq Q^*]} \Delta_\sigma(\beta, \sigma, \theta, \hat{Q}_S)dF. \\ &\quad - \int C_\sigma(\sigma, \theta, Q^*)dF - 1 \end{aligned}$$

Because we already know that for $\beta = \beta^*$ the expected joint surplus is uniquely maximized at σ^*, we will study the function

$$\tilde{s}(\sigma) := s(\beta^*, \sigma) - \left(\int W(\beta^*, \sigma, \theta, Q^*(\beta^*, \sigma, \theta))dF - \sigma \right).$$

which has derivative

$$\tilde{s}'(\sigma) = -(1-\gamma)\int_{[Q^* > \bar{q}]} \Delta_\sigma(\beta^*, \sigma, \theta, \bar{q})dF - \int_{[\hat{Q}_S \geq Q^*]} \Delta_\sigma(\beta^*, \sigma, \theta, \hat{Q}_S)dF. \tag{3.7}$$

By exploiting $C_{\sigma q} \leq 0$, it is straightforward to see that $\Delta_\sigma(\beta^*, \sigma, \theta, q)$ is weakly decreasing in q, and that the first term in $\tilde{s}'(\sigma)$ is negative and the second is positive (if they do not vanish). The first term is the derivative of what ER call the "hold-up tax", this term is responsible for any potential underinvestment, and the second term is the derivative of the seller's "breach subsidy", this term may create overinvestment.

Now, in order to prove the lemma, consider first $\bar{q} = q_L$. In this case, for all $\sigma \geq \sigma^*$, since Q^* is nondecreasing in σ, the event $[Q^* > q_L]$ is equal to Θ and

$$\tilde{s}'(\sigma) = -(1-\gamma) \int \Delta_\sigma(\beta^*, \sigma, \theta, q_L) dF \leq 0.$$

Hence, $\tilde{s}$ is a monotonically decreasing function in this range. All $\sigma > \sigma^*$ then lead to a lower payoff than σ^*, hence $\max \sigma_S(q_L, p) \leq \sigma^*$. For a contract over q_H the first term in $\tilde{s}'$ vanishes for $\sigma \leq \sigma^*$, i.e., $\tilde{s}$ is a weakly increasing function. Therefore, at q_H all $\sigma < \sigma^*$ are dominated by σ^*, and $\min \sigma_S(q_H, p) \geq \sigma^*$. Finally, consider q_H and a low price p_L. By definition of p_L it holds that $\hat{Q}_S(\sigma, \theta) \leq Q^*(\beta^*, \sigma, \theta)$ for all $\theta \in \Theta$ and σ. Therefore, the function $\tilde{s}$ is weakly decreasing for $\sigma \geq \sigma^*$, hence $\sigma_S(q_H, p_L) = \{\sigma^*\}$. For the buyer, the corresponding claims follow from the assumption that $V_{\beta q} \geq 0$.

Proof of Lemma 3.2.
Again, we prove the claim only for the seller. First, let us state the required conditions more precisely. For each θ, whenever $Q^*(\beta^*, \sigma, \theta) \leq \hat{Q}_S(\sigma, \theta)$ we need that $S(\sigma, \theta, \hat{Q}_S)$ is concave in σ, i.e.,

$$C_{\sigma\sigma}(\sigma, \theta, \hat{Q}_S) - \frac{C_{q\sigma}(\sigma, \theta, \hat{Q}_S)^2}{C_{qq}(\sigma, \theta, \hat{Q}_S)} \geq 0.$$

This condition follows from Assumption 3, because the determinant of the Hessian matrix of $(\sigma, q) \mapsto C(\sigma, \theta, q)$ is positive at $q = \hat{Q}_S$. One can see here why a linear cost function might be a problem: as C_{qq} becomes small, this condition becomes harder to fulfill. Furthermore we need the condition that $W(\beta^*, \sigma, \theta, Q^*)$ is concave, meaning that

$$C_{\sigma\sigma}(\sigma, \theta, Q^*) + \frac{C_{\sigma q}(\sigma, \theta, Q^*)^2}{W_{qq}(\beta^*, \sigma, \theta, Q^*)} \geq 0,$$

which also follows from Assumption 3. Last, we need the condition $C_{\sigma\sigma}(\sigma, \theta, \bar{q}) \geq 0$, which is also implied by convexity of C in both variables.

Since s is continuous in q, p and σ (which is straightforward to check), according to Berge's theorem, the argmax correspondence $\sigma_S(q, p)$ is upper hemicontinuous. Since upper hemicontinuity coincides with continuity if the correspondences are functions, for Assumption 2 to hold it suffices that the function $\sigma \mapsto s(\sigma, \beta^*)$ has a unique maximizer for all q and p. We therefore show that s is strictly concave, given that Assumption 3 holds. For this we need that the derivative (see equation 3.6) is decreasing in σ. It suffices to show that the continuous integrand is piecewise decreasing, which can be done by calculating the piecewise derivatives and using the above conditions.

Proof of Proposition 3.3.
Since because of Assumption 2 the best responses have a continuous selection, we may assume that $\sigma_S(q,p)$ and $\beta_B(q,p)$ are continuous functions. For all $p \in [p_L, p_H]$, define

$$\bar{q}_S(p) := \{q \in [0, q_H] : \ \sigma_S(q,p) = \sigma^*\}$$

and

$$\bar{q}_B(p) := \{q \in [0, q_H] : \ \beta_B(q,p) = \beta^*\}.$$

¿¿From Lemma 3.1 and the intermediate value theorem it follows that these sets are nonempty for each p. Since the derivative $s'(\beta^*, \sigma^*)$ (equation (3.6) is weakly increasing in q, these sets must also be convex, i.e., $\bar{q}_S$ and $\bar{q}_B$ are compact and convex valued upperhemicontinuous correspondences. Consider first the case that they are functions.[11] Lemma 3.1 tells us that $\bar{q}_S(p_L) = q_H \geq \bar{q}_B(p_L)$ and $\bar{q}_B(p_H) = q_H \geq \bar{q}_S(p_H)$. Applying the intermediate value theorem again yields existence of a $\bar{p}$ such that $q_S(\bar{p}) = q_B(\bar{p}) =: \bar{q}$. This contract $(\bar{q}, \bar{p})$ thus leads to β^* as a best response to σ^* and σ^* as a best response to β^*.

If the correspondences $\bar{q}_S$ and $\bar{q}_B$ are not single-valued, their graphs are still pathwise connected and a similar argument applies: Since $\bar{q}_S$ and $\bar{q}_B$ are compact and convex valued upperhemicontinuous correspondences, the same is true for $d := \bar{q}_S - \bar{q}_B$. We have to show that there exists a $\bar{p}$ with $0 \in d(\bar{p})$. We know that $d(p_L)$ contains nonnegative elements, therefore we can define $\bar{p} = \max\{p \in [p_L, p_H] : \ d(p) \cap [0, q_H] \neq \emptyset\}$. Then we can take any sequence $(p_n)_n \subset [\bar{p}, p_H]$ with limit $\bar{p}$. For the limit $\bar{d} := \lim_n d(p_n)$ we know that both $\bar{d} \in d(\bar{p})$ and $\bar{d} \leq 0$. Convexity of $d(\bar{p})$ then implies that $0 \in d(\bar{p})$.

Proof of Corollary 1.
The derivative of $\tilde{s}(\sigma)$, as calculated in the proof of Lemma 3.1 (equation 3.7), evaluated at σ^*, must vanish at the optimal contract. The corollary follows since for the kind of functions defined in Assumption 4 it holds that

$$\Delta_\sigma(q) = -C_1'(\sigma)(Q^* - q) \text{ and } \Delta_\beta(q) = V_1'(\beta)(Q^* - q). \tag{3.8}$$

Proof of Proposition 3.5.
When the price is p_L, the buyer makes a profit on each unit, i.e., $\hat{Q}_B = \bar{q}$ for all θ. When price is p_H, it holds that $\hat{Q}_S = \bar{q}$ for all θ. Expected payoff is analogous to the case with an intermediate price and can be rearranged to look as follows (again only for the seller):

[11] This holds for example if the inequalities $C_{\sigma q} \leq 0$ and $V_{\beta q} \geq 0$ hold strictly everywhere, Q^* is continuous in θ, $\gamma \in (0,1)$, and σ^* and β^* are interior solutions.

$$s(\sigma,\beta) = \int W(\beta,\sigma,\theta,Q^*)dF - \sigma - \int B(\beta,\theta,\bar{q})dF \tag{3.9}$$
$$-(1-\gamma)\int_{[Q^*>\bar{q}]} \Delta(\beta,\sigma,\theta,\bar{q})dF - (1-\lambda)\int_{[Q^*\leq\bar{q}]} \Delta(\beta,\sigma,\theta,\bar{q})dF$$

with $p = \lambda p_L + (1-\lambda)p_H$. The claim can now be proved following the same steps as in the proof of Proposition 3.3, the role of the price being played by λ.

Proof of Proposition 3.6.
We prove this result independently of previous results in this chapter, because it holds without Assumption 2, and would hold also for arbitrary investment decisions and linear functions. For $\lambda = \gamma$, the seller's expected payoff functions as stated in equation (3.9) equals

$$s(\beta,\sigma) = \bar{p}\bar{q} + (1-\gamma)\left(\int -C(\sigma,\theta,\bar{q})dF - \sigma\right) \tag{3.10}$$
$$+\gamma\left(\int W(\beta,\sigma,\theta,Q^*)dF - \sigma\right) - \gamma\int V(\beta,\theta,\bar{q})dF \tag{3.11}$$

with $\bar{p} = \gamma p_L + (1-\gamma)p_H$. In this case, the payoff functions are identical to the ones that result from specific performance in ER. Next, consider the defining equation of σ^*, which is that for all other σ

$$\int W(\beta^*,\sigma^*,\theta,Q^*(\sigma^*,\beta^*,\theta))dF - \sigma^* \geq \int W(\beta^*,\sigma,\theta,Q^*(\sigma,\beta^*,\theta))dF - \sigma. \tag{3.12}$$

Furthermore, from the definition of Q^* we know that

$$W(\beta^*,\sigma,\theta,Q^*(\sigma,\beta^*,\theta)) \geq W(\beta^*,\sigma,\theta,Q^*(\sigma^*,\beta^*,\theta)) \quad \text{for all } \sigma,\theta. \tag{3.13}$$

From these two equations, it follows that

$$\sigma^* \in \arg\max_{\sigma} \int -C(\sigma,\theta,Q^*(\sigma^*,\beta^*,\theta))dF - \sigma \tag{3.14}$$

Since we assumed the special payoff functions defined in Assumption 4 it follows that with $\bar{q} = \int Q^*(\beta^*,\sigma^*,\theta)dF$

$$\sigma^* \in \arg\max_{\sigma} \int -C(\sigma,\theta,\bar{q})dF - \sigma. \tag{3.15}$$

Hence, when $\beta = \beta^*$, all terms in the seller's payoff function are maximized at σ^*, and it is straightforward to show that the same holds symmetrically for the buyer.

Proof of Proposition 3.7.
The derivative of the seller's payoff function, evaluated at σ^*, is

$$-(1-\gamma)\int_{[Q^*\geq\hat{Q}_B]}\Delta_\sigma(\sigma^*,\theta,\hat{Q}_B)d\pi.$$

Therefore, a necessary condition for first best investment levels is $\gamma = 1$ or $Q^*(\sigma^*,\beta^*) \leq \hat{Q}_B(\beta^*)$ almost surely. In case of $\gamma = 1$, choose $\bar{p}$ such that

$$\int V_\beta(\beta^*,\theta,\hat{Q}_B)d\pi = 1$$

at $p = \bar{p}$. Then choice of β^* is a dominant strategy for the buyer, and σ^* is the seller's best response. If $Q^*(\sigma^*,\beta^*) \leq \hat{Q}_B(\beta^*)$ a.s., the buyer will overinvest except if $Q^*(\sigma^*,\beta^*) = \hat{Q}_B(\beta^*)$ a.s., which would lead to investments σ^* and β^* and efficient trade without renegotiation. However, for this to hold the price function must equal the cost function, which therefore has to be deterministic and linear.

3.8 Examples

In this appendix we compute two examples, to explore for which type of functions Assumption 2 is likely to hold. In the first example P^* is deterministic, such that the concavity assumption becomes very important. The second example shows that the first best can also sometimes be reached although the cost function is linear, as long as there is enough variance in P^*. Let $\gamma = \frac{1}{2}$ and

$$\begin{aligned} C(\sigma,\theta,q) &= \frac{1}{2\sigma}q + c\frac{q^2}{2\theta}, \\ V(\beta,\theta,q) &= \left(\frac{4}{3}c + \frac{7}{3} - \frac{1}{2\beta}\right)q - \frac{q^2}{2\theta}. \end{aligned}$$

In the specification of the model the investment cost was normalized to be linear, but it can as well be any convex function. For this example, we take $\sigma^2/2$ to be the cost of investment σ. The uncertainty parameter θ is assumed to be uniformly distributed on the interval $[1,2]$. The efficient quantity is

$$Q^*(\beta,\sigma,\theta) = \left(\frac{4}{3}c + \frac{7}{3} - \frac{1}{2\beta} - \frac{1}{2\sigma}\right)\frac{\theta}{1+c}.$$

Calculations reveal that $\sigma^* = \beta^* = 1$. Since the equilibrium price

$$P^*(\beta, \sigma, \theta) = \frac{c}{1+c}\left(\frac{4}{3}c + \frac{7}{3} - \frac{1}{2\beta} - \frac{1}{2\sigma}\right) + \frac{1}{2\sigma}.$$

does not depend on θ, the only candidate for an efficient contract is $\bar{q} = \frac{8}{3}$ and $\bar{p} = \frac{4}{3}c + \frac{1}{2}$. The sufficient condition in Assumption 3 is fulfilled if $c > 3/16$.[12] For very low c, this contract leads to a saddle point instead of a maximum of the seller's payoff function at σ^*. This can be seen by calculating the second derivative for $\sigma \geq \sigma^*$: as c goes to zero, it becomes positive.

This example is one in which, once investment is sunk, only one party breaches the contract. Nevertheless, since the overinvesting party faces hold-up and non-breach contingencies, the equilibrium is efficient if the payoff functions are sufficiently concave. As the cost function approaches a linear and deterministic one, the first best ceases to be attainable.

This does not necessarily hold if there is a random element in the linear term, such that always both parties face the risk of breaching. Consider the following variant of the preceding example:

$$C(\sigma, \theta, q) = \left(\frac{1}{2\sigma} + \theta_1\right) q$$

$$V(\beta, \theta, q) = \left(\frac{7}{3} - \frac{1}{2\beta} + \theta_1\right) q - \frac{q^2}{2\theta_2}.$$

That is, we set $c = 0$ and to the contingency we add a new component which makes marginal cost volatile. The part θ_1 is a market shock which affects both the buyer's valuation and the seller's cost (which could be opportunity cost). The part θ_2 only affects the buyer, and is again uniformly distributed on $[1, 2]$. With regard to θ_1, we assume that it is uniformly distributed on $[0, 1]$. The efficient quantity is now

$$Q^*(\beta, \sigma, \theta) = \left(\frac{7}{3} - \frac{1}{2\beta} - \frac{1}{2\sigma}\right) \theta_2.$$

Looking for the optimal contract, we get the following equation from the seller's maximization problem:

$$\int_{[Q^* \leq \bar{q}]} \left(\bar{p} - \frac{1}{2}\right) (\bar{q} - Q^*)\, d\theta_2 = \int_{[Q^* > \bar{q}]} \frac{1}{2} (Q^* - \bar{q})\, d\theta_2$$

One obvious solution is $\bar{q} = 2$ and $\bar{p} = 1$. All solutions are characterized by

$$\bar{p}_S(q) = \frac{1}{2} + \frac{(\frac{3}{4}\bar{q} - 2)^2}{2(\frac{3}{4}\bar{q} - 1)^2}$$

[12] This bound is even lower if the convex investment cost is taken into account.

for all $q_H = \frac{8}{3} > \bar{q} > q_L = \frac{4}{3}$. The buyer's payoff fulfills all assumptions. Unfortunately, the condition that characterizes the optimal contract for the buyer becomes quite complex. As numerical solutions of the two equations we get $\bar{q} = 2.039$ and $\bar{p} = 0.8956$.

4

Signaling an outside option

4.1 Introduction

While so-called general investments are rewarded by the market, specific investment loses a large part of its value if it is used outside a particular relationship.[1] If specific investments are not contractible, their level will depend on other characteristics of the relationship, for example on the ownership right to the asset that will be improved by the investment. If there are no other contracts, bargaining power and alternative use of the investment determine the investor's share in the investment's returns. For example, an employee may increase her human capital in the safe knowledge that it cannot be taken away from her and has a value for other employers as well. Subcontractors that produce an input for a downstream firm will have some incentives to innovate if they are granted ownership of the asset that they work to improve, or a legal title to the innovation that they develop. Nevertheless, if there is a large discrepancy between the asset's value in the current relationship and the next best alternative, investment incentives might be diluted for fear of opportunistic behavior of the other party.

Many investments in machines or human capital are a mix of specific and general investment. In fact, how specific an investment is also depends on the

[1] In the terminology of Klein, Crawford and Alchian (1978), general investments create no appropriable rents, while the quasi-rents that are generated by specific investments may be subject to opportunistic behavior of the other party in the relationship. Another context in which this distinction between general and specific investment is important is the acquisition of human capital (Becker (1964)). The worker, who owns the property right to his human capital, will always acquire an efficient amount of general investment, but firm-specific captial is not so easily induced.

characteristics of the investing party, e.g. on its access to the market for the asset or its ability to transform the asset to general use. For a worker who has the entrepreneurial ability to use his training to start his own business, all training might be considered general. In contrast, for a worker who has to rely on finding a job in a similar business the specificity of the investment depends on his cost of switching jobs. The degree of specificity determines the investing party's ex post bargaining position, but it may in fact be a hidden characteristic of the investing party. In this chapter, we explore the consequences of assuming that the investing party from the outset has private information about the outside option.

The baseline model that we use in this chapter is a simplified version of the property rights model developed by Grossman and Hart (1986) and Hart and Moore (1990)[2]. An upstream supplier invests into an asset, which is specific to the relationship with a downstream buyer. It is not possible to write detailed long-term contracts, instead the buyer later makes a take-it-or-leave-it offer to the seller and thus determines how they share the return to investment. If the buyer knows what the seller can maximally accomplish without the cooperation of the buyer, this is what the buyer will offer to the seller. The seller's return and incentives to invest are then completely determined by this outside option. For the seller to have any incentive to invest at all in this setting, he must own the asset.

We make the assumption that the degree of specificity, as captured by the best alternative use of the asset, is private information to the seller. In this situation, the downstream party tries to deduce the upstream firm's outside option from the level of relationship-specific investment. If the seller is very reluctant to invest, the buyer infers that the seller probably fears a hold-up because he cannot use the investment elsewhere. The buyer will then indeed make a low offer. In contrast, if the seller is very eager to invest, the buyer may conclude that his private value from the investment is high, hence she has to make a high offer. Now of course the possibility arises that the seller with a low outside option mimics the high type and invests more. This effect could potentially mitigate the hold-up problem and lead to higher investment.

We find that this game, in which the seller tries to signal a high outside option with his investment, has a unique equilibrium (modulo out-of-equilibrium beliefs and strategies). If the seller's outside option is known to be relatively low compared to the value of the investment to the buyer, all types of sellers invest the same amount. They choose the same investment as the type with the maximum outside option does under symmetric information. Clearly, in

[2] See also Hart (1995), Farrell and Gibbons (1995) and de Meza and Lockwood (1998).

such a pooling equilibrium investments and joint surplus are higher than in the case with complete information.

In general, the equilibrium is a hybrid, or semi-pooling, equilibrium. There is a cut-off type such that all sellers with a lower outside option pool on this type's strategy. This cut-off type, and all higher ones, mix between their own and all higher types' complete information investments. All these types hence separate in the sense that they choose different strategies. Because of the randomization, however, a chosen investment does not give away the type ex post. An observed investment could have been chosen by any type who would invest less under complete information. While the information asymmetry leads to higher investment, this effect is traded against the inefficiency generated by the non-investing party trying to appropriate part of the information rents. How the joint surplus compares to the case with complete information therefore depends on the parameters of the model.

That in our model relationship-specific investment can be used as a signal for an outside option distinguishes this chapter from the rest of the literature. The idea that private information about outside options can lead to rents that foster investment has been addressed before, eg. by Malcomson (1997) and Sloof (2008). In these papers, the outside options are realized after investment decisions have been made. Although there is no signaling going on, such models yield similar qualitative predictions: in comparison to the standard hold-up model, which excludes all ex post frictions and focuses on inefficient preparations, there are now greater inefficiencies ex post and less ex ante. In particular, investment levels can be too high relative to their later use. A characteristic of the signaling model, in comparison, is a "bluffing" element that leads to an equilibrium in mixed strategies.[3]

Signaling models by now have a long tradition in economics, starting with Spence (1973), who models education as a wasteful signal of productivity.[4] It is possible to reveal private information with signals like for example warranties or high prices as signals for quality, because the cost of the signal differs across types. In contrast, in our model the cost of investment depends only indirectly on types. Because all types of sellers have the same cost of investment, types only matter if the other party uses her bargaining power and makes low offers. In particular, types of sellers completely separate under symmetric information, while in the original Spence model the wasteful signal is not used at all under symmetric information. This also means that by definition signaling cannot lead to underinvestment in that model, but this

[3] This outcome of an equilibrium in mixed strategies due to a commitment problem is reminiscent of equilibria in hold-up problems with asymmetric information as studied in Gul (2001) and Gonzales (2004).

[4] For an excellent survey of signaling and screening models, see Riley (2001).

changes if one allows education to be productive (see Weiss (1983)). More related to the present chapter is recent work on signaling that assumes productive investment and shows that signaling leads to higher investment and even to a Pareto improvement. This includes Hermalin (1998), in which a leader may signal a worthwhile project by exerting high effort, and Daughety and Reinganum (2009), in which a signaling motive helps a team to overcome a free-riding problem.

The remainder of the chapter is organized as follows. In Section 4.2, the outside option signaling game is introduced. We first solve it for a finite type space in Section 4.3. In Section 4.4 we analyze the case that the type space is a continuum. We analyze both these cases because it is much more natural to think about the problem using a finite type space, but the solution has a more tractable from in the limit of a continuous type space. We also discuss how changes in the timing or information structure would change the outcome of the game; in particular we analyze a version with commitment in Section 4.5. Proofs not given in the text can be found at the end of the chapter.

4.2 The outside option signaling game

The model describes an interaction between a downstream buyer-manufacturer and an upstream supplier who has to tailor his production processes to the needs of the buyer.[5] In the game with complete information, the seller chooses an investment $i \in \mathrm{I}$, at cost $c(i)$, to improve the value of an asset/good to be traded. If seller and buyer work together, they can generate a value of $v(i)$, while the value of the good or asset to the seller without the buyer is only the fraction $\theta v(i), \theta \in \Theta \subset [0, 1]$.[6] The buyer observes the investment and the value of the asset and makes an offer about how to share the surplus with the seller. If the seller rejects the offer, he gets $\theta v(i)$ from taking his outside option, while the buyer is left with zero. If the seller accepts, they split the generated surplus as proposed by the buyer.

ASSUMPTION 5 *We assume that* $\mathrm{I} = \mathbb{R}$, *that the functions* v *and* c *are differentiable, increasing, and concave resp. strictly convex. Furthermore* $v(0) \geq 0$, $c(0) = 0$, $c'(0) = 0$, *and* $\lim_{i \to \infty} c'(i) = \infty$.

The buyer has no way to commit to a particular reaction or to write a contract that conditions on i or $v(i)$ or that specifies a particular bargaining

5 As explained in the introduction, the model is very abstract and therefore fits a variety of settings, including an employer-employee relationship.

6 There does not need to be a deterministic relationship between the value and the investment. As long as the downstream party can observe the investment and the value, the analysis remains valid if $v(i)$ represents the expected value.

game. Instead she makes a take-it-or-leave-it offer to the seller, which is optimal for her from an ex post perspective, but not necessarily from an ex-ante perspective. If θ is the type of the buyer, i the seller's investment, $o \in [0,1]$ the buyer's offer, expressed as a share of the surplus, and $a \in \{0,1\}$ the acceptance decision of the seller, then the seller's payoff is given by

$$(ao + (1-a)\theta)v(i) - c(i)$$

and the buyer's payoff by

$$a(1-o)v(i).$$

This game can be easily solved by backward induction. The seller will accept all offers $o > \theta$, and since the buyer can always offer a little bit more, we assume that the seller (except maybe if $\theta = 1$) accepts all offers $o \geq \theta$. The buyer will offer a share θ of the realized surplus, which the seller will accept, leaving him a profit of $\theta v(i) - c(i)$ from investment i. In anticipation of this return to investment the seller invests

$$i^c(\theta) = \arg\max \theta v(i) - c(i),$$

which given our assumptions always exists and is unique. Therefore also the inverse of i^c exists, which we denote by $\theta^c : i^c(\Theta) \to \Theta$. The seller's payoff under complete information, in dependence on the outside option θ, is denoted by

$$u^c(\theta) = \max_i \theta v(i) - c(i).$$

Note that the derivative of u^c is equal to $v \circ i^c$, and in particular, u^c is increasing and strictly convex. [7]

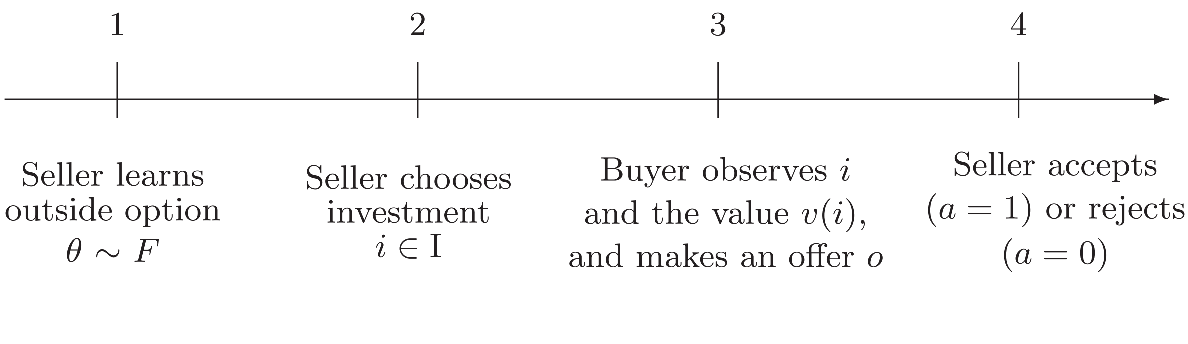

Fig. 4.1. Timeline of the outside option signaling game.

[7] We could alternatively make this, or other conditions from which it follows, our assumption.

In the game with incomplete information, θ is private information of the upstream seller. The sequence of events is illustrated in Figure 4.1. We assume that first the seller learns his type θ, which is drawn from a type space $\Theta \subset [0, 1]$ according to a distribution function F. The buyer only knows the distribution of the outside option, but not the realized value. She observes the seller's investment, forms beliefs about the outside option and then makes a take-it-or-leave-it offer that is optimal for her given her updated beliefs about the acceptance threshold of the seller. We are interested in perfect Bayesian equilibria of this game, and in such an equilibrium a seller of type θ will accept an offer if and only if it is greater than the outside option. We therefore take this acceptance decision, the same as in the game with perfect information, for granted, and deal with the following payoff functions: if the seller is of type θ and invests i, and the buyer makes an offer o, then the seller gets $\max(\theta, o)v(i) - c(i)$ and the buyer gets $(1 - o)v(i)$ if $\theta \leq o,$ and 0 else.

A strategy of the seller specifies an investment for each type, possibly using a randomization device to mix over a set of investments. A strategy of the seller is a function $Q : \Theta \times \mathrm{I} \to [0, 1]$ such that $Q(.|\theta) := Q(\theta, .)$ is the distribution of investments that a type θ chooses. A strategy for the buyer maps all possible investments into a share of the surplus that she offers to the seller, where she as well may randomize over a set of offers. We write a strategy of the buyer as a function $P : \mathrm{I}\times[0, 1] \to [0, 1]$, where $P_i(o) := P(i, o)$ is the probability that the buyer's offer, when observing investment i, is less or equal to o.

If the buyer's strategy is given by P, the seller's expected profit from choosing investment i is

$$U(P, i, \theta) = \int \max(\theta, o)dP_i(o)\; v(i) - c(i),$$

and given a strategy Q of the seller, the buyer's expected payoff from the pure strategy $o : \mathrm{I} \to [0, 1]$ is

$$V(Q, o) = \int \int_{[\theta \leq o(i)]} (1 - o(i))v(i)dQ(i|\theta)dF(\theta).$$

4.3 Finite type space

In this section, we assume that $\Theta = \{\theta_1, ..., \theta_H\}$ with $0 \leq \theta_1 < \theta_2 < ... < \theta_H < 1$.[8] We shortcut $i^c(\theta_k) =: i_k$. In the following, we fix a perfect Bayesian

[8] The assumption $\theta_H < 1$ is made only for simplicity. We could easily add types $\theta \geq 1$ who would always invest $i^c(\theta)$ and get no acceptable offer from the buyer.

equilibrium of the signaling game (P, Q). We will derive properties of (P, Q), in order to eventually arrive at a characterization of all equilibra of the outside option signaling game. Let I^* be the set of investments that are chosen with positive probability in the equilibrium (P, Q), and let $\Theta^*(i)$ denote the set of all types that choose $i \in \mathrm{I}^*$ with positive probability. We denote the equilibrium payoff to a seller of type θ by $u^*(\theta)$, i.e., with this notation we have for all $i \in \mathrm{I}^*$ and $\theta \in \Theta^*(i)$ that $u^*(\theta) = U(P, i, \theta)$.

Note that $u^*(\theta) \geq u^c(\theta)$, because a type θ can always guarantee himself the payoff $u^c(\theta)$ independent of the buyer, by investing $i^c(\theta)$ and taking his outside option. Similarly, because the seller's payoff is weakly increasing in θ for all offers and investments, $U(P, i, \theta)$ and $u^*(\theta)$ are weakly increasing in θ. A higher type could always play a lower type's strategy and get at least the same payoff as that type.

In the following, we will first show that if an investment i may occur at all in equilibrium, then it is chosen with positive probability by the type $\theta^c(i)$ that chooses i under symmetric information, and by none of the higher types. Then, in Lemma 4.1, we show that investing i is optimal for all types from θ_1 to $\theta^c(i)$. Finally, in Prop. 4.3 we will answer the question which investments will be chosen in equilibrium. The reader who is not interested in the proofs may skip the lemmas leading to Prop. 4.3 which contains the main result of this section.

When the buyer observes an investment $i \in \mathrm{I}^*$, she updates that the seller must have an outside option in $\Theta^*(i)$. The share she offers will therefore also lie in $\Theta^*(i) \subset \{\theta_1, ..., \theta_H\}$, and it will never be more than the highest possible type would accept, i.e. the offer is not higher than $\theta_m = \max \Theta^*(i)$. The profit to type θ_m from choosing i is therefore equal to $\theta_m v(i) - c(i)$, which would be strictly smaller than $u^c(\theta_m)$ if $i \neq i_m$. Therefore $i = i_m$, which means that if an investment i occurs in the signaling equilibrium, then $\theta^c(i)$ is the highest type to choose this investment. In particular, only investments $i_k, k = 1, ..., H$ occur in equilibrium.

We will sometimes use the one-to-one relationship between θ_k and i_k and express everything in types. We can also identify the buyer's offer with the type that just accepts it, and then write the equilibrium strategies P and Q as matrices. An entry p_{kl} of P stands for the probability of offer θ_l when investment i_k is observed, and an entry q_{kl} in Q is the probability of type k investing i_l, or "mimicking" type l. Since we have shown that in any equilibrium the mixed strategy of type θ_k has support $\{i_k, ..., i_H\}$ and the buyer's random offer following investment i_k takes on values in $\{\theta_1, ..., \theta_k\}$, equilib-

---

That is, a type $\theta \geq 1$ seller would neither mimic other types nor be mimicked himself.

rium strategies P and Q are triangular matrices. Equilibrium conditions for strategies (P,Q) in matrix form then look as follows:

(i) $q_{kl} > 0$ implies that

$$l \in \arg\max_{m} \sum_{j=1}^{m} p_{mj} \max(\theta_j, \theta_k) v(i_m) - c(i_m),$$

(ii) for each l with $i_l \in \mathrm{I}^*$, $p_{lj} > 0$ implies that

$$j \in \arg\max_{k} (1-\theta_k) \sum_{j=1}^{k} f_j q_{jl}.$$

We will show next that the set of best responses to P of a given type θ_k includes all investments that are greater or equal than i_k and are chosen at all in the equilibrium. In other words, if an investment i_k is chosen at all, then it is optimal for every type smaller or equal to the corresponding type θ_k.

Lemma 4.1. *For all $i_k \in \mathrm{I}^*$ it holds that $U(P, i_k, \theta) = u^*(\theta)$ for all $\theta = \theta_1, ..., \theta_k$.*

Proof. We know already that $U(P, i_k, \theta_k) = u^*(\theta_k)$. First, we show that this also holds for the lowest type, i.e. that $U(P, i_k, \theta_1) = u^*(\theta_1)$. To this end, let θ_l be the lowest type with this property, i.e., $U(P, i_k, \theta_l) = u^*(\theta_l)$ and $U(P, i_k, \theta) < u^*(\theta)$ for all $\theta < \theta_l$. Since no type below θ_l chooses i_k, the offer following it cannot be lower than θ_l. Type l's expected payoff then does not depend on him being type θ_l, but every lower type would get the same payoff when investing i_k :

$$U(P, i_k, \theta_l) = \int o dP_{i_k}(o) v(i_k) - c(i_k) = U(P, i_k, \theta) \text{ for all } \theta \leq \theta_l. \tag{4.1}$$

Payoff monotonicity then implies that $U(P, i_k, \theta) = u^*(\theta)$ for any type $\theta \leq \theta_l$, hence $l = 1$.

Second, we show that for a seller of type θ_l the investments that are best responses to P can be found by maximizing $P_i(\theta_{l-1})v(i)$ over all $i \in \mathrm{I}^*$, where we define $P_i(\theta_0) = 0$. More precisely, the claim is that for all $l = 1, ..., H$

$$\arg\max_{i \in \mathrm{I}^*} U(P, i, \theta_l) = \arg\max_{i \in \mathrm{I}^*} P_i(\theta_{l-1}) v(i) \subset \arg\max_{i \in \mathrm{I}^*} U(P, i, \theta_{l-1}).$$

It is clear that the claim implies the lemma, since it tells us that

$$i_k \in \arg\max_{i \in \mathrm{I}^*} U(P, i, \theta_k) \subset ... \subset \arg\max_{i \in \mathrm{I}^*} U(P, i, \theta_1).$$

It remains to prove the claim, which we will do by induction. Since we know that $U(P, i, \theta_1) = u^*(\theta_1)$ for all $i \in \mathrm{I}^*$, it holds for $l = 1$ for the appropriate definitions. Assume the claim is true for type $l - 1 \geq 1$. For all $i \in \mathrm{I}^*$ with $u^*(\theta_{l-1}) = U(P, i, \theta_{l-1})$ type θ_l's payoff is

$$U(P, i, \theta_l) = u^*(\theta_{l-1}) + (\theta_l - \theta_{l-1})P_i(\theta_{l-1})v(i). \tag{4.2}$$

while for any $i' \in \mathrm{I}^*$ with $U(P, i', \theta_{l-1}) < u^*(\theta_{l-1})$ it holds that

$$U(P, i', \theta_l) < u^*(\theta_{l-1}) + (\theta_l - \theta_{l-1})P_{i'}(\theta_{l-1})v(i'). \tag{4.3}$$

Using the induction hypothesis, we have that for any such i and i'

$$P_{i'}(\theta_{l-1})v(i') = P_{i'}(\theta_{l-2})v(i') \leq P_i(\theta_{l-2})v(i) \leq P_i(\theta_{l-1})v(i),$$

hence we have shown that $U(P, i', \theta_l) < U(P, i, \theta_l)$. The remainder of the claim follows easily.

To summarize, we have shown so far that in any equilibrium, while there may be investments that do not occur at all, every investment that does occur is chosen by the type that would invest the same amount with symmetric information. Furthermore, all lower types' payoff from choosing this investment equals their equilibrium payoff. In order to be consistent with this structure, the buyer's strategy must induce all these indifferences. This observation gives rise to the following lemma.

Lemma 4.2. *For all k and $i \in \mathrm{I}^*$ with $i > i_k$ it holds that*

$$P_i(\theta_k)v(i) = \frac{u^*(\theta_{k+1}) - u^*(\theta_k)}{\theta_{k+1} - \theta_k}. \tag{4.4}$$

Moreover, for all $i_m, i_k \in \mathrm{I}^*$ with $m \geq k$ it holds that $p_{mk} > 0$.

Proof. The first claim follows from the proof of Lemma 4.1, because there we had that for all $i \in \mathrm{I}^*$ with $i > i_k$ it holds that

$$u^*(\theta_{k+1}) = u^*(\theta_k) + (\theta_{k+1} - \theta_k)P_i(\theta_k)v(i).$$

To show the last claim of the lemma, note first that for any type θ_k with $i_k \in \mathrm{I}^*$ it must be true that $p_{kk} > 0$, because else $U(P, i_k, \theta_{k-1})$ is too low: if $p_{kk} = 0$, this payoff is equal to

$$((1 - p_{kk})\theta_{k-1} + p_{kk}\theta_k)v(i_k) - c(i_k) = \theta_{k-1}v(i_k) - c(i_k) < u^c(\theta_{k-1}).$$

Second, assume that for $m > k$ as in the lemma we have $p_{mk} = 0$. Then

$$0 = P_{i_m}(\theta_k)v(i_m) - P_{i_m}(\theta_{k-1})v(i_m) = \frac{u^*(\theta_{k+1}) - u^c(\theta_k)}{\theta_{k+1} - \theta_k} - \frac{u^c(\theta_k) - u^*(\theta_{k-1})}{\theta_k - \theta_{k-1}},$$

whence

$$u^c(\theta_k) = u^*(\theta_{k+1})\frac{\theta_k - \theta_{k-1}}{\theta_{k+1} - \theta_{k-1}} + u^*(\theta_{k-1})\frac{\theta_{k+1} - \theta_k}{\theta_{k+1} - \theta_{k-1}}.$$

As mentioned before, the function u^c is strictly convex. Therefore, and because

$$\theta_k = \theta_{k+1}\frac{\theta_k - \theta_{k-1}}{\theta_{k+1} - \theta_{k-1}} + \theta_{k-1}\frac{\theta_{k+1} - \theta_k}{\theta_{k+1} - \theta_{k-1}},$$

we have that

$$u^c(\theta_k) < u^c(\theta_{k+1})\frac{\theta_k - \theta_{k-1}}{\theta_{k+1} - \theta_{k-1}} + u^c(\theta_{k-1})\frac{\theta_{k+1} - \theta_k}{\theta_{k+1} - \theta_{k-1}}.$$

Hence, $p_{mk} > 0$.

Now that we have some idea about the offers that the buyer must be willing to make, we turn to a description of the buyer's behavior, in order to pin down the seller's equilibrium strategy. The details can be found in the proof of the following proposition that describes the structure of an equilibrium. But first we need more notation and an assumption:

ASSUMPTION 6 *Let* $R(\theta) := (1-\theta)F(\theta)$ *and* $\bar{k} := \min\{k : R(\theta_k) > R(\theta_{k+1})\}$.[9] *We assume that* R *is strictly concave on* $\{\theta_{\bar{k}}, ..., \theta_H\}$.

Assume for a moment that all types choose the same investment i. Then $R(\theta)$ describes the buyer's expected share of the surplus $v(i)$ if she makes a take it or leave it offer of θ. The maximum $\bar{\theta}$ of this function is the offer that she would make in a pooling equilibrium. Can a pooling equilibrium exist? Since the highest type θ_H chooses i_H in any equilibrium, if all types pool on the same investment, this must be i_H. It follows that there is such a pooling equilibrium if and only if $\bar{\theta} = \theta_H$. This suggests that complete pooling is only possible for types lower than $\bar{\theta}$, and since a separating type could easily be mimicked by a lower type, equilibria must typically be of a hybrid form and involve mixed strategies.

Proposition 4.3. *If Assumption 6 holds, then an equilibrium of the signaling game must have the following form: No investment below* $i_{\bar{k}}$ *is chosen. A type* θ_k *with* $k \geq \bar{k}$ *mixes between all investments in* $\{i_k, ..., i_H\}$, *with expected payoff equal to* $u^c(\theta_k)$. *All types* θ_k *with* $k \leq \bar{k}$ *mix over* $\{i_{\bar{k}}, ..., i_H\}$ *with payoff* $u^c(\theta_{\bar{k}})$. *When observing investment* i_k, *the buyer mixes between offers in* $\{\theta_{\bar{k}}, ..., \theta_k\}$, *and her expected payoff from any such offer is* $(1-\theta_k)v(i_k)$.

[9] Let $\theta_{H+1} = 1$.

Proof. See Section 4.7 at the end of the chapter.

All equilibria of the outside option signaling game lead to the same expected payoffs. Refinements to pin down beliefs following zero probability events are not needed for this result. The reason is that even if an investment i_k does not trigger an acceptable offer from the buyer, this investment would still allow type θ_k to get $u^c(\theta_k)$ by himself.

From all the indifference conditions that have to be met in an equilibrium we are able to obtain an equilibrium candidate. Combining Prop. 4.3 and Lemma 4.2 yields for all $k \geq \bar{k}$ and $m > k$

$$P_{i_m}(\theta_k) = \frac{u^c(\theta_{k+1}) - u^c(\theta_k)}{(\theta_{k+1} - \theta_k)v(i_m)} \quad \text{and} \quad P_{i_k}(\theta_k) = 1, \tag{4.5}$$

as well as for $k < \bar{k}$

$$P_{i_m}(\theta_k) = 0. \tag{4.6}$$

The equilibrium conditions for the seller's strategy are

$$(1 - \theta_l) \sum_{j=1}^{l} f_j q_{jk} = (1 - \theta_k) \sum_{j=1}^{k} f_j q_{jk} \quad \text{for all } k \geq l \geq \bar{k} \tag{4.7}$$

and

$$(1 - \theta_l) \sum_{j=1}^{l} f_j q_{jk} \leq (1 - \theta_k) \sum_{j=1}^{k} f_j q_{jk} \quad \text{for all } l < \bar{k}. \tag{4.8}$$

Due to the definition of $\bar{k}$, the latter condition can be fulfilled by defining

$$q_{jk} = q_{\bar{k}k} \quad \text{for all } j < \bar{k}. \tag{4.9}$$

Let us further define $\lambda_k := \frac{f_k(1-\theta_k)(1-\theta_{k-1})}{\theta_k - \theta_{k-1}}$ and $\lambda_{H+1} := 0$. Possible values for the q_{jk} are:

$$q_{\bar{k}k} = \frac{\lambda_k - \lambda_{k+1}}{R(\theta_{\bar{k}})} \quad \text{for all } k > \bar{k} \tag{4.10}$$

$$q_{\bar{k}\bar{k}} = 1 - \frac{\lambda_{\bar{k}+1}}{R(\theta_{\bar{k}})} \tag{4.11}$$

$$q_{jk} = \frac{\lambda_k - \lambda_{k+1}}{\lambda_j} \quad \text{for all } k \geq j > \bar{k} \tag{4.12}$$

Proposition 4.4. *The strategies described in equations (4.5), (4.6), (4.9), (4.10), (4.11) and (4.12) form an equilibrium of the outside option signaling game.*

Proof. See Section 4.7 at the end of the chapter.

Example

We look at an example with three types to illustrate the different kinds of equilibrium and the uniqueness issue. First, since $R(\theta)$ is the buyer's expected share of $v(i)$ if all types choose the same investment i and the buyer offers θ, pooling on the investment i_3 is an equilibrium if and only if $(1-\theta_3) = \max_\theta R(\theta)$. We write this equilibrium in the matrix form described at the beginning of this section:

$$\mathrm{Q} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \ \mathrm{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that beliefs out of equilibrium, i.e. after observing an investment $i \neq i_3$, are not pinned down uniquely. Consequently also the first two rows in P are not uniquely determined.

In case $(1-\theta_2)F(\theta_2) = \max_\theta (1-\theta)F(\theta)$ an equilibrium is of the following form:

$$\mathrm{Q} = \begin{pmatrix} 0 & q_{12} & 1-q_{12} \\ 0 & q_{22} & 1-q_{22} \\ 0 & 0 & 1 \end{pmatrix}, \ \mathrm{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & p_{32} & 1-p_{32} \end{pmatrix}$$

Again, the first row of P does not have to be the unit vector. To see how the structure of Q translates into the condition for R, let $\mu_2 := q_{22}f_2 + q_{12}f_1$ be the probability of i_2 being chosen (here the same as the probability of any lower investment being chosen). The conditions for the buyer are

- $(1-\theta_3)(1-\mu_2) = (1-\theta_2)(F(\theta_2)-\mu_2)$ which is equivalent to $\mu_2 = \frac{R(\theta_2)-R(\theta_3)}{\theta_3-\theta_2}$. This expression is always less or equal to 1, and it is non-negative iff $R(\theta_2) \geq R(\theta_3)$.
- $(1-\theta_2)(F(\theta_2)-\mu_2) \geq (1-\theta_1)(F(\theta_1)-q_{12}f_1)$ which is equivalent to $q_{12}f_1 \geq \frac{R(\theta_1)-R(\theta_2)}{1-\theta_1} + \frac{(1-\theta_2)\mu_2}{(1-\theta_1)}$
- $(1-\theta_2)\mu_2 \geq (1-\theta_1)q_{12}f_1$ which is equivalent to $q_{12}f_1 \leq \frac{(1-\theta_2)}{(1-\theta_1)}\mu_2$

Obviously, the last two conditions can only be fulfilled if $R(\theta_1) \leq R(\theta_2)$. If this holds, the solutions are $q_{12} = \frac{(1-\theta_2)\mu_2}{R(\theta_1)} - \Delta$ for any $0 \leq \Delta \leq \frac{R(\theta_2)-R(\theta_1)}{R(\theta_1)}$. Thus, in this case the solution is typically not unique. If we make the restriction $q_{12} = q_{22}$, the last two conditions, which state that the buyer prefers offering θ_2 to offering θ_1, read

- $q_{12}f_1 \geq \frac{R(\theta_1)-R(\theta_2)}{1-\theta_1} + \frac{R(\theta_2)q_{12}}{(1-\theta_1)} \iff 1 \geq q_{12}$
- $f_1 \leq \frac{(1-\theta_2)}{(1-\theta_1)}F(\theta_2) \iff R(\theta_1) \leq R(\theta_2)$

That is, we immediately have a solution, given by $q_{12} = q_{22} = \frac{R(\theta_2)-R(\theta_3)}{F(\theta_2)(\theta_3-\theta_2)}$. This is not surprising, because here the pooling condition (R increasing) holds up to θ_2. The proposed equilibrium in Prop. 4.4 also uses this fact. The buyer's expected profit does not depend on the values of q_{12} and q_{22}, only on μ_2.

If $(1-\theta_1)F(\theta_1) = \max_\theta(1-\theta)F(\theta)$, then the equilibrium is unique:

$$\mathrm{Q} = \begin{pmatrix} q_{11} & q_{12} & 1-q_{11}-q_{12} \\ 0 & q_{22} & 1-q_{22} \\ 0 & 0 & 1 \end{pmatrix}, \mathrm{P} = \begin{pmatrix} 1 & 0 & 0 \\ p_{21} & 1-p_{21} & 0 \\ p_{31} & p_{32} & 1-p_{31}-p_{32} \end{pmatrix}$$

For the values of the entries, see Proposition 4.4. The expressions may become complex, that is why we look at a continuous strategy space in the next section.

We know from Prop. 4.3 that a strategy of the form

$$\mathrm{Q} = \begin{pmatrix} q_{11} & 0 & 1-q_{11} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

cannot be part of an equilibrium. This can be checked explicitly here, showing that for this to be an equilibrium it must be true that $R(\theta_1) = \max_\theta R(\theta)$ and R convex, contradicting our assumption that R is concave. While it might be possible to relax this assumption and still say something about the resulting equilibria, we do not address this question in this chapter.

4.4 Continuous type space.

The expressions for equilibrium strategies will have a simpler form in this section, which treats the continuous type space as the limit case. Hence, in this section $\Theta = [\theta_L, \theta_H]$. We assume that F is an atomless distribution on Θ with density $f > 0$, for which the derivative f' exists.

ASSUMPTION 7 *F is log-concave.*

Analogous to the previous section, we define $\bar{\theta} = \theta_H$ if $R'(\theta) \geq 0$ on Θ, and else

$$\bar{\theta} = \inf\{\theta \in \Theta : \ R'(\theta) < 0\}, \tag{4.13}$$

and have

Lemma 4.5. *Given Assumption 7, R is concave on $[\bar{\theta}, \theta_H]$ and $\bar{\theta} = \arg\max_\theta R(\theta)$.*

Proof. To show that F log-concave (which is implied by f log-concave) is sufficient for this property of R, we will show first that

$$R''(\theta) \geq 0 \Rightarrow R'(\theta) > 0.$$

The second derivative of R is

$$R''(\theta) = (1-\theta)f'(\theta) - 2f(\theta)$$

such that $R''(\theta) \geq 0$ implies that $f'(\theta) \geq 0$ and

$$(1-\theta) \geq \frac{2f(\theta)}{f'(\theta)}.$$

Hence,

$$R'(\theta) = (1-\theta)f(\theta) - F(\theta) \geq \frac{2f(\theta)^2 - F(\theta)f'(\theta)}{f'(\theta)} \geq \frac{f(\theta)^2}{f'(\theta)} > 0.$$

From the definition of $\bar{\theta}$, where we have a local maximum, the claim easily follows.

Proposition 4.6. *Given Assumption 7, an equilibrium of the signaling game is given by*

$$P_i(\theta) = \begin{cases} 0 & \theta < \bar{\theta} \\ \frac{v(i^c(\theta))}{v(i)} & \bar{\theta} \leq \theta \leq \theta^c(i) \\ 1 & \theta \geq \theta^c(i) \end{cases} \tag{4.14}$$

and $Q(i|\theta) = Q(i|\bar{\theta})$ *for all* $\theta < \bar{\theta}$, *and for all* $\theta \geq \bar{\theta}$

$$Q(i|\theta) = \begin{cases} 0 & i < i^c(\theta) \\ 1 - \frac{(1-\theta^c(i))^2 f(\theta^c(i))}{(1-\theta)^2 f(\theta)} & i^c(\theta) \leq i < i^c(\theta_H) \\ 1 & i = i^c(\theta_H) \end{cases} \tag{4.15}$$

The proof is straightforward and therefore omitted. It can also be shown that this equilibrium is the limit of the equilibrium found in the previous section (Prop. 4.4) as the partition becomes finer.

4.4.1 Surplus Comparison

In the following paragraphs, we compare different timings and information regimes with respect to the payoff that is generated for the seller and the buyer as well as the joint surplus. In some applications as for example the mobility of a worker, it seems realistic that the worker knows his mobility but the employer never learns it until it is too late. Alternatively, it may be the case that the worker learns his outside option only after making the firm-specific investment. In a market setting, it may be that the outside option is known to both sides from the start, or that both sides learn it after investment decisions have been made. We always evaluate payoffs and surplus with respect to the distribution F, and that is also how the expectations in the following expressions should be understood.

First we look at the case of complete information. In this case the outside option is common knowledge even before investment is undertaken. The seller's expected profit is $E[u^c(\theta)]$ and the buyer gets $E[(1-\theta)v(i^c(\theta))]$. The expected joint surplus is $E[S(i^c(\theta))]$ with $S := v - c$.

If the outside option becomes common knowledge only after the investment is sunk, and is not known before to any party, the expected social surplus is $S(i^c(E[\theta]))$. If we assume that $S(i^c(\theta))$ is a concave function in θ (eg. $v''' - c''' \leq 0$) then this surplus is higher than under complete information. The seller gets $u^c(E[\theta])$ and is therefore worse off than in the complete info case, because he cannot prepare for his outside option. The buyer is better off with $(1-E[\theta])v(i^c(E[\theta]))$, capturing the quasi-rent from low types who invest too much.

A third timing and information structure of the game is that the seller, and only the seller, learns the outside option later. In this case, there is no signaling motive. The buyer makes an offer of $\bar{\theta}$ and the seller invests $i^c(E[\theta \vee \bar{\theta}])$. While the investment is higher than in the two cases above, it is not always put to its best use, as all types above $\bar{\theta}$ reject the offer. The seller gets $u^c(E[\bar{\theta} \vee \theta])$ which is more than in the previous case, as he enjoys some informations rents. The buyer gets $R(\bar{\theta})v(i^c(E[\theta \vee \bar{\theta}]))$.

Finally, in the signaling equilibrium (Prop. 4.6), a seller with outside option θ gets $\max(u^c(\theta), u^c(\bar{\theta}))$, i.e. the seller's expected profit is

$$F(\bar{\theta})u^c(\bar{\theta}) + \int_{\bar{\theta}}^{\theta_H} u^c(\theta)f(\theta)d\theta.$$

To find the buyer's surplus in the signaling equilibrium, note first that $-R''(\hat{\theta})$ is the probability density of investment on $[\bar{\theta}, \theta_H)$. Therefore, the buyer's expected payoff is

$$\int_{\bar{\theta}}^{\theta_H} -R''(\theta)(1-\theta)v(i^c(\theta))d\theta + (1-\theta_H)^2 f(\theta_H)v(i(\theta_H))$$

We see that the seller has an incentive to learn the outside option early, because in the signaling equilibrium his expected payoff is $E[u^c(\theta \vee \bar{\theta})] > u^c(E[\theta \vee \bar{\theta}])$. It is also better for him in expected terms if it is known that he is aware of his outside option, while its value stays secret. In the next section we consider the case that the investment decision is contractible. It is clear that in that case, lower seller types are unambiguously better for the buyer.

4.5 The case with commitment

In the game that is studied in the main part of this chapter, all the buyer can do is make a take it or leave it offer based on her updated beliefs. In this section we shall explore the consequences of full commitment and ask what would happen if the buyer could offer a binding contract conditional on investment before the seller moves. We assume that she still cannot observe the seller's type, and characterize the optimal screening contract. While in the signaling model the seller moves first, now the buyer can act before the seller takes the investment decision. [10]

Proposition 4.7. *If the buyer can write a contract on the investment decision, the outcome involves investment of* $i^c(1)$ *and inefficient separation if* $\theta \geq \bar{\theta}$*: Such a type* θ *takes the outside option with probability* $p(\theta) = \frac{v(i^c(\theta))}{v(i^c(1))}$. *Each seller type is left with the same payoff as in the case without commitment,* $\max(u^c(\theta), u^c(\hat{\theta}))$. *The buyer gets* $S(i^c(1)) - u^c(\bar{\theta}) - \int_{\bar{\theta}}^{\theta_H} S(i^c(\theta))dF(\theta)$.

Proof. We use the revelation principle and let a general contract be a map from types into outcomes that satisfies the incentive compatibility constraints of each type of seller telling the truth. The buyer also has to take into account that the seller can go for his outside option, then getting a payoff of $\theta v(i)$ after having invested an amount i, or $u^c(\theta)$ ex ante.

All that matters for truth telling and participation of the seller is his expected payoff, and the buyer in addition cares for the surplus created by the contract. Therefore, it is sufficient to concentrate on contracts of the form $(t(\theta), i(\theta), p(\theta))$, where $t(\theta)$ is an up-front payment from the seller to the buyer, $i(\theta)$ is the investment that an announced type θ is required to make, and $p(\theta)$ is the probability of separation. With probability $1 - p(\theta)$, buyer and seller collaborate and the seller gets the whole ex post surplus $v(i(\theta))$. There is no loss of generality in assuming this form of contracts, because all payoff transfers from the seller to the buyer can be handled by the up-front payment $t(\theta)$. Given such a contract, the expected payoff to a seller of type θ who pretends to be of type $\tilde{\theta}$ is

[10] Adverse selection problems with type-dependent reservation utilities have been addressed before, but in different frameworks (Moore (1985), Jullien (2000)). Our problem has a much simpler structure, but is not a special case of these results.

$$(1 - p(\tilde{\theta}))v(i(\tilde{\theta})) + p(\tilde{\theta})\theta v(i(\tilde{\theta})) - c(i(\tilde{\theta})) - t(\tilde{\theta}).$$

A truth-telling seller creates the joint surplus $S(i(\theta)) - p(\theta)(1 - \theta)v(i(\theta))$, and gets $u_S(\theta) = S(i(\theta)) - p(\theta)(1 - \theta)v(i(\theta)) - t(\theta)$ for himself. The buyer's optimization problem is the following:

$$\max \int_{\theta_L}^{\theta_H} t(y)dF(y),$$

subject to the incentive compatibility constraint

$$u_S(\theta) \geq u_S(\tilde{\theta}) + (\theta - \tilde{\theta})p(\tilde{\theta})v(i(\tilde{\theta})) \tag{IC}$$

and the ex ante participation constraint

$$u_S(\theta) \geq u^c(\theta), \tag{PC}$$

which have to hold for all $\theta, \tilde{\theta} \in [\theta_L, \theta_H]$.

It may seem intuitive that an optimal contract specifies efficient investment, because the seller types do not differ with respect to the cost of investment, only with respect to the outside option. The screening device therefore is the probability of separation, not the investment. However, since in order to separate the seller's types this probability must be positive, it is not obvious that $i(\theta) = i^c(1)$, because $i^c(1)$ is not the optimal preparation for every type (which would be $i^c(1-(1-\theta)p(\theta))$). In particular, so far the formulation also allows for some types not participating and choosing $p = 1, i = i^c(\theta), t = 0$.

To see that setting $i(\theta) = i^c(1)$ is without loss of generality, consider any contract $(t(\theta), i(\theta), p(\theta))$. The contract $(\tilde{t}(\theta), \tilde{i}(\theta), \tilde{p}(\theta))$ defined by

$$\tilde{t}(\theta) = t(\theta) + S(i^c(1)) - S(i(\theta) \geq t(\theta),$$

$\tilde{i}(\theta) = i^c(1)$, and

$$\tilde{p}(\theta) = p(\theta)\frac{v(i(\theta))}{v(i^c(1))} \in [0, 1]$$

leads to the same IC and PC constraints and weakly higher expected profit for the buyer. In particular, this means that excluding types is not a good idea for the buyer.

For any $p : [\theta_L, \theta_H] \to [0, 1]$ that is part of an IC contract, if $p(\tilde{\theta}) = 0$ for some type $\tilde{\theta}$, then we know that lower types pool on this type, i.e. $u_S(\tilde{\theta}) = u_S(\theta)$ for all types $\theta \leq \tilde{\theta}$. In the buyer's optimal contract it will then hold that $p(\theta) = 0$ and $t(\theta) = S(i^c(1)) - u^c(\tilde{\theta})$ for all $\theta \leq \tilde{\theta}$. We therefore now take a threshold $\theta^0 \in \Theta$ as given and replace the IC constraints by the requirement that p is nondecreasing and

$$u_S(\theta) = \int_{\theta_0}^{\theta} p(y)v(i^c(1))dy + u^c(\theta^0).$$

We define $P^0 := \{p : [\theta^0, \theta_H] \to (0,1], \text{ nondecreasing}\}$. Following the standard method of finding an optimal screening contract we write the problem as

$$\max_{p \in P^0} S(i^c(1)) - u^c(\theta^0) - \int_{\theta^0}^{\theta_H} (R'(y)+1)p(y)v(i^c(1))dy$$
$$\text{s.t.} \int_{\theta^0}^{\theta} p(y)v(i^c(1))dy \geq u^c(\theta) - u^c(\theta^0).$$

Because $R'(\theta)+1 \geq 0$, $p(\theta)$ must be as small as possible. This suggests that the PC should bind everywhere, which we will indeed show next. First, because the objective function can also be written as

$$S(i^c(1)) - u_S(\theta_H) - \int_{\theta^0}^{\theta_H} R'(y)p(y)v(i^c(1))dy$$

it is clear that $\theta^0 \geq \bar{\theta}$. Furthermore, for the part that depends on p we can use integration by parts to get

$$\begin{aligned} &u_S(\theta_H) + \int_{\theta^0}^{\theta_H} R'(y)p(y)v(i^c(1))dy \\ =&(1-\theta_H)f(\theta_H)u_S(\theta_H) - R'(\theta^0)u^c(\theta^0) - \int_{\theta^0}^{\theta_H} R''(y)u_S(y)dy \\ \geq&(1-\theta_H)f(\theta_H)u^c(\theta_H) - R'(\theta^0)u^c(\theta^0) - \int_{\theta^0}^{\theta_H} R''(y)u^c(y)dy \end{aligned}$$

This shows that the objective function is maximized if the PC is binding everwhere. For this to be true, the buyer would have to set

$$p(\theta) = \frac{v(i^c(\theta))}{v(i^c(1))},$$

which is indeed increasing, hence must be the solution to the optimization problem. Finally, we find the optimal θ_0: Solving

$$\max_{\theta_0} S(i^c(1)) - u^c(\theta_0) - \int_{\theta_0}^{\theta_H} (R'(\theta)+1)v(i^c(\theta))d\theta$$

yields $\bar{\theta}$ as the optimal cut-off value.

The optimal contract induces higher investment now that i is verifiable, as there should be no hold-up problem. In fact, that the buyer offers no contract for any investment other than $i^c(1)$ means that there is overinvestment relative to the investment's later use. The buyer promises a contract over the full surplus $v(i^c(1))$ with some probability, in exchange for an up-front payment. The seller can choose among a menu of contracts consisting of combinations of separation probabilities and up-front payments

$$\left(\frac{v(i)}{v(i^c(1))}, S(i^c(1)) - S(i)\right), \quad i \in [i^c(\bar{\theta}), i^c(\theta_H)],$$

or trade for sure and pay $S(i^c(1)) - u^c(\bar{\theta})$ up-front. This contract excludes higher types with positive probability, which would be impossible without a form of commitment.

4.6 Conclusion

We introduced private information about the reservation value in a simple property rights model.[11] The simplicity of the model allowed us to fully characterize the resulting equilibrium payoffs, which are uniquely determined. The equilibrium involves pooling up to a certain type of outside option, such that all lower types get the same payoff. Because they accept all offers in equilibrium these types are not distinguishable, even ex post. Higher types follow a mixed strategy and on average obtain the same payoff as with complete information. The seller has to randomize since there is a strong force against a separating equilibrium in this model: if only high types choose a certain investment and get high offers, they will be mimicked by lower types.

In the outside option signaling game, there is a gap between the chosen investment and the investment that would result if the seller obtained the full return to his investment. We have shown that this gap vanishes if investment is verifiable. The gap would also shrink if the seller had greater bargaining power than in the game that was analyzed. For example, if the bargaining game was modeled as the seller making a take it or leave it offer with probability α and

[11] In fact, the model is essentially an ultimatum game with a prior investment stage, in which the responder invests to increase the pie that he and the proposer can share. The model can then be interpreted as the responder having private information about the payoff he gets when rejecting an unfair offer. Simply obtaining a proportion of the pie as payoff is of course not a good model of human behavior, and therefore for this application the reader is referred to more realistic models of behavior like the game in von Siemens (2007), which then leads to a more complex signaling structure.

the buyer only with probability $1 - \alpha$, then a higher α would increase the surplus and the seller's payoff. Since there is more investment on average, the buyer's payoff is non-monotonic in α. It would also be interesting to allow for more complex bargaining games at the ex post stage. One game that should leave the results unchanged obtains if the buyer can make repeated offers; but if both players can make offers, results will change and become difficult to derive (cf. Skryzpazc (2004)).

There are a couple of other extensions of the model that present themselves. One interesting task for future work is to allow the payoff that the buyer gets when the seller takes the outside option to be dependent on the seller's type. This would admit a greater set of applications, in particular the interpretation of the outside option as suing the buyer for payment, with private information about the probability of winning.[12] Another possible extension is the case of pure rent-seeking, in which the investment increases the outside value but is of little use inside the relationship. Investment can still be used as a signal for profitable outside opportunities, but higher investment is no longer more efficient.

[12] See Chonné and Linnemer (2008) for a related model in the context of pretrial bargaining and investment in trial preparation.

4.7 Proofs

Proof of Proposition 4.3.
Let $i_k \in \mathrm{I}^*$. When observing i_k, the buyer's expected profit from offering θ_l is $G(\theta_l|i_k)(1-\theta_l)$, where

$$G(\theta_l|i_k) = \frac{\sum_{j=1}^{l} f_j q_{jk}}{\sum_{j=1}^{k} f_j q_{jk}}.$$

We know from Lemma 4.2 that to be consistent with the seller's behavior, the buyer, when observing i_k, has to offer all θ_j, $i_j \in \mathrm{I}^*, j \leq k$ with positive probability. She will offer θ_k if

$$\sum_{j=1}^{k} f_j q_{jk}(1-\theta_k) \geq \sum_{j=1}^{l} f_j q_{jk}(1-\theta_l) \text{ for all } l,$$

and θ_l if

$$\sum_{j=1}^{k} f_j q_{jk}(1-\theta_k) = \sum_{j=1}^{l} f_j q_{jk}(1-\theta_l).$$

As a first step, we write down all inequalities that define the buyer's behavior in an equilibrium (P,Q). Denote by

$$K := \{k : \; i_k \in \mathrm{I}^* \backslash \{i_H\}\}$$

all chosen investments that are strictly smaller than i_H. We treat H separately because we have to account for the fact that Q is a stochastic matrix, i.e., that the row entries add up to one. For all $j,l \leq k, l,k \in K$ the following inequalites must hold:

$$\begin{aligned} \sum_{i=1}^{j} f_i(\theta_k-\theta_j)q_{ik} + \sum_{i=j+1}^{k} f_i(\theta_k-1)q_{ik} &\leq 0 \\ -\left(\sum_{i=1}^{l} f_i(\theta_k-\theta_l)q_{ik} + \sum_{i=l+1}^{k} f_i(\theta_k-1)q_{ik}\right) &\leq 0 \\ -q_{jk} &\leq 0 \end{aligned}$$

plus (straightforward calculation) for all $l < H, i \in K$

$$R(\theta_H) - R(\theta_l) \geq \sum_{j=1}^{l} \sum_{j \leq k \in K} f_j(\theta_l - \theta_H) q_{jk} + \sum_{j=l+1}^{H-1} \sum_{k \in K} f_j(1 - \theta_H) q_{jk}$$

$$R(\theta_i) - R(\theta_H) \geq \sum_{j=1}^{i} \sum_{j \leq k \in K} f_j(\theta_H - \theta_i) q_{jk} + \sum_{j=k+1}^{H-1} \sum_{j \leq k \in K} f_j(\theta_H - 1) q_{jk}$$

$$1 \geq \sum_{j \leq l \in K} q_{ji}$$

We are going to treat the variables we are looking for as one big vector, denoted by q. That is, the entries in q are indexed by $jk, 1 \leq j \leq k, k \in K$. Similarly, we define a vector μ^{jk} by $\mu^{jk}_{ik} = f_i(\theta_k - \theta_j)$ for all $i \leq j$ and $\mu^{jk}_{ik} = f_i(\theta_k - 1)$ for all $i > j$ and zero else. Furthermore, define a vector μ^l by $\mu^l_{jk} = f_j(\theta_l - \theta_H)$ for all $j \leq l$ and $\mu^l_{jk} = f_j(1 - \theta_H)$ for all $j > l$. Last, let 1^j denote a vector with $1^j_{jk} = 1$ for $j \leq k \in K$ and 0 else. And let e^{jk} be a vector with $e^{jk}_{jk} = 1$ and 0 else.

Our inequalities now read

$$\begin{aligned} -e^{jk} q &\leq 0 && 1 \leq j \leq k, k \in K \\ 1^j q &\leq 1 && j = 1, ..., H-1 \\ \mu^{jk} q &\leq 0 \text{ for all } k \in K, j < k \text{ and} \geq 0 \text{ for } j \in K \\ \mu^l q &\leq R(\theta_H) - R(\theta_l) \text{ for all } l < H \text{ and } \geq 0 \text{ for } l \in K \end{aligned}$$

As the second step, we find a system of inequalities that is an alternative of this system, i.e. that has a solution if and only if this one has none. We use Theorem 22.1 of Rockafellar (1970) to get the following alternative system:

(i) $\sum_{j=1}^{H-1} \beta_j + \sum_{l=1}^{H-1} \delta_l (R(\theta_H) - R(\theta_l)) < 0$
(ii) $\sum_{j=1}^{H-1} 1^j \beta_j + \sum_{jk} \mu^{jk} \gamma_{jk} + \sum_{l=1}^{H-1} \mu^l \delta_l \geq 0$

where we are looking for coefficients $\beta_j \geq 0$, $j = 1, ..H-1$, γ_{jk} (≥ 0 if $j \notin K$), δ_l, (≥ 0 if $l \notin K$). For the analysis, it is convenient to write the second equation as an equation in each coefficient jk with $k \in K$ and $j \leq k$

$$\beta_j + \sum_{i=1}^{j-1} \gamma_{ik} f_j(\theta_k - 1) + \sum_{i=j}^{k-1} \gamma_{ik} f_j(\theta_k - \theta_i) + \sum_{l=1}^{j-1} \delta_l f_j(1 - \theta_H) + \sum_{l=j}^{H-1} \delta_l f_j(\theta_l - \theta_H) \geq 0$$

Let $\hat{k} = \min K$. We claim that $\bar{k} = \hat{k}$ and first show that $R(\theta_l) \leq R(\theta_{\hat{k}})$ for $l < \hat{k}$. Assume not. Then there is a solution with $\delta_l = \gamma_{lk} = 1$ and

$\delta_{\hat{k}} = \gamma_{\hat{k}k} = -1$ and all other coefficients equal to zero: The first inequality is obviously satisfied, and for the second, since $k \geq \hat{k} > l$ always holds, there are only three cases to distinguish, $j > \hat{k}$, $l < j \leq \hat{k}$, and $j \leq l$.

Similarly, one can show that $R(\theta_{\hat{k}+1}) \leq R(\theta_{\hat{k}})$ is also necessary, because else there is a solution with $\delta_{\hat{k}+1} = \gamma_{\hat{k}+1k} = 1$ and $\delta_{\hat{k}} = \gamma_{\hat{k}k} = -1$. The easy case distinctions are again left to the reader. Hence, $\hat{k} = \bar{k}$. Note that we could have shown more generally that $K \subset \{k \text{ with } R(\theta_k) \geq R(\theta_{k+1})\}$.

Next we show that K is an interval. Assume to the contrary that there is a gap in K, i.e, there exist $l < m < h$ with $m \notin K$, $l = \max\{k \in K, k \leq m\}$ and $h = \min\{k \in K, k \geq m\}$. There is a $\lambda \in (0,1)$ with $(1-\lambda)\theta_h + \lambda\theta_l = \theta_m$. Define $\delta_l = \gamma_{lk} = -\lambda, \delta_m = \gamma_{mk} = 1, \delta_h = \gamma_{hk} = -(1-\lambda)$ for all relevant $k \in K$. Then the first condition holds because R is concave on K: $\lambda R(\theta_l) + (1-\lambda)R(\theta_h) - R(\theta_m) < 0$. That the second condition always holds with equality is seen immediately if $k \leq l$, for which this condition takes the form $\theta_m - \theta_H - \lambda(\theta_h - \theta_H) - (1-\lambda)(\theta_l - \theta_H) = 0$. For the remaining case $k \geq h$ there has to be again a case distinction regarding j, each case leading to the same result. Thus concavity of R implies that there are no gaps in chosen investment, $K = \{\bar{k}, ..., H-1\}$.

Proof of Proposition 4.4.
First, we check that the strategies fulfill equation 4.7:

$$
\begin{aligned}
(1-\theta_l)\sum_{j=1}^{l} f_j q_{jk} &= (1-\theta_l)\left(\sum_{j=1}^{\bar{k}} f_j \frac{\lambda_k - \lambda_{k+1}}{R(\theta_{\bar{k}})} + \sum_{j=\bar{k}+1}^{l} \frac{\lambda_k - \lambda_{k+1}}{\lambda_j}\right) \\
&= (1-\theta_l)\left(\frac{\lambda_k - \lambda_{k+1}}{1-\theta_{\bar{k}}} + \sum_{j=\bar{k}+1}^{l}\left(\frac{\lambda_k - \lambda_{k+1}}{1-\theta_j} - \frac{\lambda_k - \lambda_{k+1}}{1-\theta_{j-1}}\right)\right) \\
&= \lambda_k - \lambda_{k+1},
\end{aligned}
$$

which is independent of l. Similarly for the remaining cases.

Next, note that

$$
\frac{R(\theta_k) - R(\theta_{k-1})}{\theta_k - \theta_{k-1}} = \frac{f_k(1-\theta_{k-1})}{(\theta_k - \theta_{k-1})} - F(\theta_k) = \frac{f_k(1-\theta_k)}{(\theta_k - \theta_{k-1})} - F(\theta_{k-1})
$$

and therefore

$$
\lambda_k - \lambda_{k+1} = (1-\theta_k)\left(\frac{R(\theta_k) - R(\theta_{k-1})}{\theta_k - \theta_{k-1}} - \frac{R(\theta_{k+1}) - R(\theta_k)}{\theta_{k+1} - \theta_k}\right) \geq 0.
$$

Also,

$$
R(\theta_{\bar{k}}) - \lambda_{\bar{k}+1} \geq 0 \Leftrightarrow R(\theta_{\bar{k}}) - R(\theta_{\bar{k}+1}) \geq 0.
$$

These conditions imply that all $q_{jk} \geq 0$. We still need to show that they add up to one:

$$\sum_{k=j}^{H} q_{jk} = \sum_{k=j}^{H} \frac{\lambda_k - \lambda_{k+1}}{\lambda_j} = 1 \qquad \text{for all } j > \bar{k}$$

$$\sum_{k=\bar{k}}^{H} q_{j\bar{k}} = 1 - \frac{\lambda_{\bar{k}+1}}{R(\theta_{\bar{k}})} + \sum_{k=\bar{k}+1}^{H} \frac{\lambda_k - \lambda_{k+1}}{R(\theta_{\bar{k}})} = 1$$

Here we have that all low types follow the same strategy. If such a restriction is not imposed, there may be more possible values for the strategies.

5

Renegotiation-proof relational contracts with side payments

5.1 Introduction

Relational contracts are self-enforcing informal agreements that arise in many long-term relationships, often in response to obstacles to write exogenously enforceable contracts. Examples include the non-contractible aspects of employment relations, illegal cartel agreements, or buyer-seller relations in which complete formal contracts are too costly to write. Agreements between countries also often have the nature of a relational contract, when there is no institution that is able or willing to enforce compliance with the agreed terms. In these examples, monetary transfers play a role in the relationships, be it in form of prices, bonuses or other compensation schemes, and could thus also be used to sustain the relational contract. Moreover, the relational contracts are drafted and negotiated, and meetings continue to take place as the relationship unfolds. In this chapter, we analyze relational contracts under these circumstances: with renegotiation and the possibility to make monetary transfers.

As an illustration how side-payments can be used in a relational contract consider the case of collusive agreements. Cartels sometimes use compensation schemes to make sure that each firm in the cartel stays with the target (see Harrington (2006) for details[1]). A cartel member that violates the agreement is required to buy a certain quantity from a competitor, or to transfer a valuable customer to a competitor. Such compensation schemes seem more robust to renegotiation than threatening with an immediate price war after a violation of the agreement. Price wars are costly for all firms, and therefore cartel members will be tempted to agree to ignore the violation. In contrast,

[1] For a list of cartels in which such compensation schemes have been used see also the introduction of Harrington and Skrzypacz (2007).

if a deviating firm must pay a fine, competitors gain from the punishment and have therefore no incentive to renegotiate the agreement. In order to induce a firm to pay the compensation there must be a threat of a real punishment in case no payment is made, a punishment that does not require the voluntary participation of the punished. Renegotiation may therefore play a role again.

The chapter investigates these issues and provides a characterization of renegotiation proof relational contracts given arbitrary discount factors. We study infinitely repeated two player games with perfect monitoring where in each period players can make monetary transfers before playing a simultaneous move game. We translate Abreu's (1988) optimal penal codes to this set-up and show that every Pareto optimal subgame perfect payoff can be achieved using a class of simple strategies, which we call *stationary contracts.* In such a stationary contract, the same action profile is played in every period, while an up-front transfer is used to achieve the required distribution of surplus. Transfers on the equilibrium path can be used to transfer slack from one player's incentive constraint to the other's.[2] A player who deviates is required to pay a fine to the other player, and after payment equilibrium play is resumed. If he does not pay, there is a single punishment action before there is another chance to make a monetary transfer and return to the equilibrium path.

In the second part of the chapter, we characterize renegotiation-proof stationary contracts, and show that again, one can often restrict the analysis to the class of stationary contracts to find payoffs that survive renegotiation-proofness refinements. The literature offers several concepts of renegotiation-proofness for infinitely repeated normal form games.We adapt strong perfection as defined in Rubinstein (1980) as well as Farell and Maskin (1989)'s concepts of weak and strong renegotiation-proof equilibria to our setting and observe that the timing of negotiations and transfers plays a role.

Since a period consists of two stages, a crucial question is at what times renegotiation is possible. We first follow Levin (2003), who studies optimal subgame perfect equilibria in a repeated principal-agent relationship, and assume that the players only meet to negotiate at the beginning of a period, before a payment is made. This assumption seems reasonable in situations where payments can be organized quickly such that whenever there is sufficient time to meet and renegotiate future actions there is also sufficient time to make side payments before future actions are conducted.

---

[2] That it suffices to look at stationary equilibrium play paths in an environment with side payments is used in many articles on relational contracting, such as Baker, Gibbons and Murphy (2002), Levin (2003), Doornik (2006), and Blonski and Spagnolo (2003).

As in Levin's framework, renegotiation-proofness has no bite with this timing; all Pareto optimal subgame perfect payoffs can be implemented in a renegotiation-proof way. In a Pareto efficient stationary contract, all continuation equilibria that start with a payment achieve the same highest possible joint payoff. At the side payment stage the punishment takes the form of the deviator paying a fine to the other player, followed by return to efficient equilibrium play. This means that if renegotiation is allowed only before transfers can be made, the threat of inefficient continuation play (which is necessary to induce payment of the fine) is never subject to renegotiation.

In the main part of this chapter, we consider a timing that also allows negotiations before the punishment for not paying is carried out. This timing is for example used in Fong and Surti (2009), who characterize subgame perfect and to some extent also renegotiation-proof outcomes in a repeated prisoner's dilemma with side payments, for all possible constellations of discount factors. The underlying assumption behind this timing is that not having made a transfer is a sunk decision, and hence the consequences can be renegotiated. This assumption makes sense in situations in which the monetary transfers can only be made at certain times, or take a long time to be organized, as should for example be the case for illegal side payments in cartels.

We first characterize strong perfect equilibrium payoffs. An equilibrium is strong perfect if all its continuation payoffs lie on the Pareto frontier of subgame perfect payoffs for that stage. The set of strong perfect equilibrium payoffs is always a subset of the Pareto frontier of subgame perfect payoffs, but it is often empty. We show that every strong perfect payoff can be achieved by a strong perfect stationary contract and derive simple conditions that allow to check for strong perfection. These conditions are used to show that in a principal agent game in which only one side has to take an action, all subgame perfect payoffs can be implemented using strong perfect stationary contracts. That in other examples strong perfect equilibria fail to exist reflects that strong perfection should be considered as a sufficient condition for renegotiation-proofness rather than a necessary requirement.

The more widely used concept of strong renegotiation proofness (SRP) only requires that continuation payoffs must lie on the Pareto frontier of weakly renegotiation proof (WRP) payoffs instead of subgame perfect payoffs. An equilibrium is WRP if players can never both improve their payoffs by negotiating from one continuation equilibrium to another. As is already known from the analysis in Baliga and Evans (2000), who study a similar setting with monetary transfers, the set of SRP payoffs converges to the Pareto frontier of individually rational stage game payoffs as the players become infinitely patient. In contrast, with intermediate discount factors SRP equilibria sometimes fail to exist. We show that if they do exist and disounting is sufficiently low[3], all SRP payoffs can be obtained by varying the up-front transfer of a single stationary contract. We also provide a simple sufficient condition for existence of SRP stationary contracts. By contrast, if discounting is too high ($\delta < \frac{1}{2}$), stationary contracts may not be sufficient to implement all Pareto optimal WRP payoffs, which can sometimes require alternation between different action profiles or money burning on the equilbrium path. We illustrate this effect in a Prisoner's Dilemma game.

In Baliga and Evans (2000), which is the most directly related work, the stage game is augmented by a side-payment mechanism such that actions and payments are chosen simultaneously. By construction, the punishment for not paying a fine is subject to renegotiation. Their analysis is hence more related to our framework than to a timing as in Levin (2003) that excludes negotiations before the stage game's actions are chosen. The difference to our framework is that we allow a larger set of discount factors as well as side payments that depend on the most recent action played. If achievable outcomes are constrained by high discounting of the future, the players have a joint incentive to use transfers that condition on past behaviour; withholding a planned transfer can already be a punishment, making defection less profitable. A characterization of optimal relational contracts for all discount factors is useful for studies that compare different regimes or institutions regarding the lowest discount factor that allows efficiency in the induced repeated game.

The chapter is organized as follows. In Section 5.2 we introduce the framework and notation. In Section 5.3 we characterize subgame perfect stationary contracts and show that all Pareto optimal subgame perfect payoffs can be implemented by a stationary contract. We show with an example how the resulting conditions can be applied to find all subgame perfect payoffs for a given discount factor. In Section 5.4 we discuss the timing of renegotiations and characterize the set of strong perfect payoffs. In Section 5.5 we show that if the future matters more than the present period ($\delta \geq \frac{1}{2}$), the WRP payoff with the highest joint payoff can be implemented by a stationary contract. If SRP payoffs exist at all, they can also be implemented using such a stationary contract. Finally, we discuss the case $\delta < \frac{1}{2}$, and then conclude in Section 5.6. Proofs have been relegated to the end of the chapter.

5.2 The game

We consider an infinitely repeated two-player game with perfect monitoring and common discount factor $\delta \in [0, 1)$. Players are indexed by $i, j \in \{1, 2\}$.

[3] We need to assume that the present period does not get a larger weight than all future periods together, i.e., that the discount factor δ does not exceed $\frac{1}{2}$.

Each period (or stage) $t = 0, 1, 2, ...$ comprises two substages, without discounting between the substages: a *side payment stage* (or *pay stage*) in which both players choose a nonnegative monetary transfer to the other player, and an *action stage* (or *play stage*) in which the players play a simultaneous move game.

The stage game is given by the continuous payoff function

$$g : A_1 \times A_2 \to \mathbb{R} \times \mathbb{R},$$

where the set A_i is the compact action space of player i. Let $A = A_1 \times A_2$, and denote the joint payoff from action profile $a = (a_1, a_2) \in A$ by $G(a) = g_1(a) + g_2(a)$. The best reply or *cheating* payoff of player i is defined by

$$c_i(a) := \max_{\{\tilde{a}:\, \tilde{a}_j = a_j\}} g_i(\tilde{a}).$$

The analysis and examples in this chapter are restricted to the case of pure strategy equilibria of the repeated game.[4] We also assume that the stage game has a Nash equilibrium in A.

In the beginning of each period, each player i may decide to make a monetary transfer to the other player, denoted $\tilde{p}_i$. We assume that the player's endowment is sufficiently large such that wealth constraints do not play a role. In order to have a compact strategy space, we require the payments to be bounded by a large bound, e.g. that payments cannot be larger than the present value of the highest possible surplus of the stage game in each period. Players are risk-neutral with quasi-linear utility. Player i's payoff in a period with net transfers $p_i := \tilde{p}_i - \tilde{p}_j$ and played action profile a is given by $g_i(a) - p_i$. We will often write payments in the form of net payments $p = (p_1, p_2)$, assuming that only the player with a positive net payment makes a monetary transfer. As is intuitively clear, simultaneous monetary transfers by both players will never be necessary to achieve a certain equilibrium payoff. [5]

A path is an infinite stream of alternating net transfers and actions, beginning with a payment or an action. Player i's average (or normalized) discounted payoff when the repeated game's outcome is given by a path $Q := \{p_t, a_t\}_{t=0}^{\infty}$ is defined as

[4] It would be possible to extend the analysis to the finite-dimensional simplex of mixed strategies and expected payoff if we assume (following Farell and Maskin (1989) and others) that a player can observe the other player's mixed strategy of a stage game and not only the realized outcome.

[5] Formally: consider a strategy profile that after some history $\{p_0, a_0, p_1, ..., a_t\}$ prescribes play of $\tilde{p}$ with $\min(\tilde{p}_1, \tilde{p}_2) = \tilde{p}_i > 0$. If $\tilde{p}_i$ is replaced by 0, and $\tilde{p}_j$ by $\tilde{p}_j - \tilde{p}_i$, then the modified strategy profile leads to the same payoffs, and is a subgame perfect equilibrium if the original has this property.

$$\tilde{u}_i(Q) := (1-\delta) \sum_{t=0}^{\infty} \delta^t (g_i(a_t) - p_{t,i}),$$

while given a path $Q := \{a_t, p_{t+1}\}_{t=0}^{\infty}$ it is

$$\tilde{u}_i(Q) := (1-\delta) \sum_{t=0}^{\infty} \delta^t (g_i(a_t) - \delta p_{t+1,i}).$$

A history that ends before stage $k \in \{pay, play\}$ in period t is a list of all transfers and actions that have occurred before stage k at that point in time. Let H^k be the set of all histories that end before stage k. A strategy σ_i of player i in the repeated game maps every history $h^{play} \in H^{play}$ into an action $\sigma_i(h^{play}) \in A_i$, and every history $h^{pay} \in H^{pay}$ into a payment $\sigma_i(h^{pay}) \geq 0$. Every strategy profile σ defines a path $Q(\sigma)$ in the usual way, and we denote $u(\sigma) = \tilde{u}(Q(\sigma))$.

Note that every history $h \in H$ defines a subgame of the repeated game. We write $\sigma|h$ for the strategy profile that starts after play according to h. While $u(\sigma) = (u_1(\sigma), u_2(\sigma))$ denotes the tuple of payoffs, we also need the total payoff $U(\sigma) := u_1(\sigma) + u_2(\sigma)$. We often make use of the fact that $u(\sigma)$ is equal to a convex combination of current period payoff (weighted by $1-\delta$) and future average payoff (weighted by δ).

Subgame perfect equilibrium means subgame perfection of the strategy profile in all of the subgames. We denote by Σ^k_{SGP} the set of subgame perfect equilibria that start in stage k. If $\sigma \in \Sigma^k_{SGP}$ is a subgame perfect equilibrium, we call $u(\sigma)$ a subgame perfect payoff. All continuation payoffs of a given equilibrium σ before a given substage $k \in \{pay, play\}$ are denoted by $\mathcal{U}^k(\sigma) = \{u(\sigma|h) : h \in H^k\}$. The set of subgame perfect payoffs is denoted by $\mathcal{U}^k_{SGP} = u(\Sigma^k_{SGP})$.

5.3 Stationary contracts and subgame perfection

5.3.1 Stationary contracts

In the following we define a class of simple stationary strategies which are helpful to characterize the Pareto-frontier of subgame perfect payoffs and to study the effects of different renegotiation-proofness requirements.

Definition 5.1. *A stationary strategy profile, denoted by $\sigma(a^e, a^1, a^2)$, is characterized by action profiles $a^e, a^1, a^2 \in A$ and a payment scheme in the following way:*

In the payment stage in period 0, there are up-front payments $p^0 = (p^0_1, p^0_2)$.

Whenever a player makes the prescribed payment in the payment stage, they play a^e *in the action stage.*

Whenever they play a^e, *or simultaneously deviate, payments* p^e *are conducted in the next payment stage.*

If player i *unilaterally deviates from the prescribed action, he pays a fine* F^i *to the other player in the subsequent payment stage.*

If ever player i *does not make the required payment, they play* a^i *in the action stage and player* i *makes an "adjustment payment"* f^i *to the other player in the subsequent payment stage.*

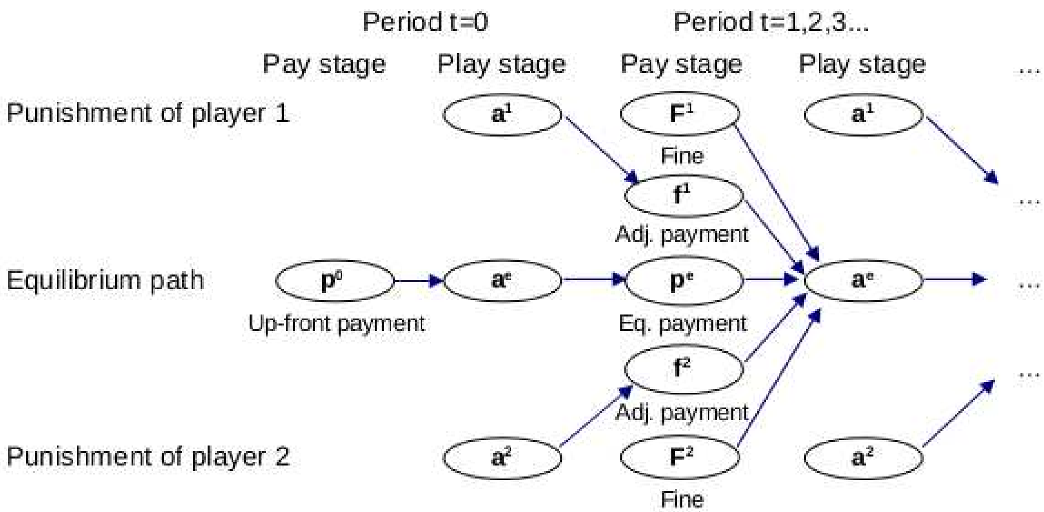

Fig. 5.1. Arrows indicate the sequence of actions according to the continuation equilibria. If player 1 (2) unilateraly deviates then the top (bottom) row will be played in the next stage. If a bilateral deviation occurs then the game continues on the equilibrium path in the next stage.

The structure of a stationary strategy profile is illustrated in Figure 5.1. To make this construction very clear, let us express a stationary strategy profile in terms of simple strategies as defined by Abreu (1988). A simple strategy profile for two players, denoted by $\sigma(Q^e, Q^1, Q^2)$, prescribes play of the initial path Q^e, while any unilateral deviation from the prescribed paths by player i is followed by play along the punishment path Q^i. In our setting, a stationary strategy profile consists of the initial path

$$Q^e = (p^0, a^e, p^e, a^e, p^e, ...)$$

and two punishment paths for player i, depending on whether the deviation occured in the side payment phase or in the action phase:

$$Q^i_{pay} = (F^i, a^e, p^e, a^e, p^e, ...)$$
$$Q^i_{play} = (a^i, f^i, a^e, p^e, a^e, p^e, ...).$$

Abreu (1988) is build around the now familiar idea that for subgame perfection the punishment does not need to fit the crime. Any deviation from a given play path can be punished by the same continuation equilibrium, namely the worst possible subgame perfect equilibrium for that player. The optimal penal codes, as such worst play paths are called in Abreu's work ((1986, 1988)), then often have a "stick and carrot" structure: they begin with the worst possible action for the punished player, and may reward him for complying with the punishment further along the path.[6] In our framework, the punishment paths have a similar structure: chosen optimally, the action a^i must have a low enough cheating payoff $c_i(a^i)$ to deter a deviation by player i. The adjustment transfer f^i is used to fine-tune the punishment. It will be positive if the punishment was not yet enough to be rewarded with equilibrium play again, and negative if the punishment was too harsh, to induce the cooperation of the punished player in the punishment.

There are two different punishment paths because a punishment can either start in the play or in the pay phase. This is due to the sequential nature of the interaction, but the intuition that the players can use a universal punishment to react to all deviations goes through in the sense that the two punishments may induce the same punishment payoff. For a given stationary strategy profile and player $i \in \{1, 2\}$, we denote this payoff by $u^i_i = \tilde{u}(Q^i_{pay})$ and define the adjustment payment f^i such that the punishment paths give the same payoff to the punished player:

$$f^i = F^i - \frac{u^i_i - g_i(a^i)}{\delta}. \tag{5.1}$$

Definition 5.2. *If a stationary strategy profile $\sigma(a^e, a^1, a^2)$ with adjustment payments given by (5.1) is a subgame perfect equilibrium, we call it a* stationary contract*.*

In the following, we find conditions that imply subgame perfection of a stationary strategy profile. It is often more convenient to think about a stationary contract in terms of the continuation payoffs it defines, and not in

[6] Such a structure with possible "repentance", where punishment does not last forever, is also useful if renegotiation proofness is an issue, which can for example be seen in Segerstrom (1988), Farell and Maskin (1989), van Damme (1989), or Farell and von Weizsäcker (2001). Return to the initial path may happen after a fixed number of punishment periods, with some probability after each period of punishment, or after the payment of a fine as in the present paper and Baliga and Evans (2000). The fine is a very powerful instrument because it additionally reduces the incentive to deviate or renegotiate for the player who collects it.

terms of the actual payments that have to be made. The equilibrium payment p^e for example plays a role for the continuation payoff on the equilibrium path starting with the action a^e, which we denote by

$$u^e = g(a^e) - \delta p^e, \tag{5.2}$$

and the fine F^i influences player i's punishment payoff

$$u_i^i = -(1-\delta)F^i + u_i^e. \tag{5.3}$$

To verify that a given stationary strategy profile is a subgame perfect equilibrium, it is sufficient to check that there are no profitable one-shot deviations. We first do this for the punishment path for player $i \in \{1, 2\}$. Irrespective of the stage in which the punishment starts, player i's payoff is always u_i^i if he complies, and either u_i^i or $c_i(a^i)(1-\delta) + \delta u_i^i$ if he deviates once and then complies again. Therefore, player i will not deviate from the punishment path if it holds that

$$u_i^i \geq c_i(a^i). \tag{5.4}$$

Since this implies that $F^i \geq f^i$, one may think of the adjustment payment f^i as a lower fine. In particular, $f^i = F^i$ holds only if player i cannot profitably deviate from a^i. Otherwise the deviator's cooperation in the punishment must be induced by the carrot of lowering the fine that is due in the next period, in fact, f^i may even be negative.

We now turn to the role of player j in player i's punishment. The condition

$$g_j(a^i)(1-\delta) + \delta((1-\delta)f^i + u_j^e) \geq c_j(a^i)(1-\delta) + \delta u_j^j$$

means that player j will play his part in a^i and that he is willing to pay $-f^i$, in case this is positive. Since the adjustment payment f^i divides the payoff on the punishment path such that player i gets u_i^i, this condition is equivalent to

$$G(a^i)(1-\delta) + \delta G(a^e) - u_i^i \geq c_j(a^i)(1-\delta) + \delta u_j^j. \tag{5.5}$$

On the equilibrium path, compliance with both a^e and the transfer p^e is implied by

$$u_i^e \geq c_i(a^e)(1-\delta) + \delta u_i^i \quad \text{for } i \in \{1, 2\}. \tag{5.6}$$

Finally, an up-front payment just like a fine divides the surplus $G(a^e)$ between the two players. Any transfer that satisfies

$$p_i^0 \leq F^i \quad \text{for } i \in \{1, 2\} \tag{5.7}$$

can be part of a stationary contract. That is, the up-front payment can be chosen to achieve any desired distribution of payoffs on the line between $(u_1^1, G(a^e) - u_1^1)$ and $(G(a^e) - u_2^2, u_2^2)$.

To summarize, we have shown that a stationary strategy profile $\sigma(a^e, a^1, a^2)$ with given transfers p^0, p^e and fines F^1, F^2 is a stationary contract if conditions (5.4), (5.5), (5.6), and (5.7) are satisfied. Since the up-front transfer can be used to achieve all possible distributions of the surplus, it is clear that the transfer on the equilibrium path only matters for the incentives not to deviate from the initial path. Therefore, we next look for conditions that allow action profiles a^e, a^1, a^2 to be part of a stationary contract given appropriate selection of the equilibrium payment. Whenever the sum of the two inequalities in (5.6) holds, then p^e can be chosen such that the individual conditions hold. Furthermore, if we are merely interested in subgame perfection, we can set the fines to the maximum level such that the cheater's continuation payoff is $u_i^i = c_i(a^i)$, $i = 1, 2$.[7] These two observations lead to the following lemma:

Lemma 5.3. *There is a stationary contract $\sigma(a^e, a^1, a^1)$ if and only if*

$$G(a^e) \geq (1-\delta)(c_1(a^e) + c_2(a^e)) + \delta(c_1(a^1) + c_2(a^2)) \tag{5.8}$$

and

$$G(a^i)(1-\delta) + \delta G(a^e) - c_i(a^i) \geq c_j(a^i)(1-\delta) + \delta c_j(a^j) \text{ for } j \neq i \in \{1, 2\}. \tag{5.9}$$

Proof. See Section 5.7 at the end of the chapter.

5.3.2 Subgame perfect payoffs

In the following we show that every Pareto optimal sugame perfect payoff can be sustained by a stationary contract. We denote the weak Pareto frontier of the set of subgame perfect payoffs by $\mathcal{P}(\mathcal{U}_{SGP}^{pay})$. Furthermore, let

$$\bar{U}_{SGP} := \sup_{u \in \mathcal{U}_{SGP}^{pay}} u_1 + u_2$$

be the maximum total payoff, and for $i = 1, 2$ let

[7] These highest fines are given by $F^i = \frac{1}{(1-\delta)}(u_i^e - c_i(a^i))$. The maximum fines become very large as the game's surplus rises. Such extreme values are not necessary, but convenient in our search for all sustainable equilibrium payoffs. There is another, equivalent, possibility to define the punishment paths, employed by Baliga and Evans (2000): We could define an equilbrium path payment $p^e = \bar{p}^i$ that makes player i's incentive constraint just binding. Fines can be much smaller if punishment of player i means paying first a fine and then switching to such a path.

$$\bar{u}^i_{SGP} := \inf_{u \in \mathcal{U}^{pay}_{SGP}} u_i$$

be the lowest payoff for player i in any subgame perfect equilibrium. Note that these values would be the same if the range of payoffs $\mathcal{U}^{pay}_{SGP}$ was replaced by $\mathcal{U}^{play}_{SGP}$, the set of subgame perfect payoffs starting with the play stage. If there is a stationary contract with equilibrium action profile $\bar{a}^e$ such that $G(\bar{a}^e) = \bar{U}_{SGP}$, then we call the stationary contract *optimal.*

Proposition 5.4. *There exists an optimal stationary contract $\sigma(\bar{a}^e, \bar{a}^1, \bar{a}^2)$ with punishment payoff $c_i(\bar{a}^i) = \bar{u}^i_{SGP}$ for player $i = 1, 2$. The Pareto frontier of subgame perfect payoffs $\mathcal{P}(\mathcal{U}^{pay}_{SGP})$ is equal to the line from $(c_1(\bar{a}^1), G(\bar{a}^e) - c_1(\bar{a}^1))$ to $(G(\bar{a}^e) - c_2(\bar{a}^2), c_2(\bar{a}^2))$.*

Proof. See Section 5.7 at the end of the chapter.

The proposition says that any Pareto optimal subgame perfect payoff can be implemented using a stationary contract $\sigma(\bar{a}^e, \bar{a}^1, \bar{a}^2)$ with maximum fines. All distributions of the optimal surplus $G(\bar{a}^e)$ that give player i at least $c_i(\bar{a}^i)$ can be achieved by changing the up-front transfer. The punishment continuation equilibrium in such a stationary contract then constitutes an optimal penal code as defined in Abreu (1988). In order to find such a stationary contract, one can use the conditions in Lemma 5.3.

5.3.3 Example (Abreu)

Abreu (1988) uses a simplified Cournot game to illustrate the superiority of optimal penal codes as compared to punishment with the worst stage game equilibrium. In the following we use this game to illustrate a simple heuristic of finding optimal subgame perfect payoffs. Two firms simultaneously choose either low (L), medium (M), or high (H) output. Stage game payoffs are given by the following matrix:

		Firm 2		
		L	M	H
	L	10,10	3,15	0,7
Firm 1	M	15,3	7,7	-4, 5
	H	7,0	5,-4	-15,-15

While total payoff is maximized if the firms choose (L,L), the unique Nash equilibrium of the stage game is (M,M); and high output minimizes the maximal payoff of the other firm. Abreu (1988) constructs optimal penal codes for the game without side payments and shows that the collusive (L, L) can be supported (in pure strategies) for $\delta \geq \frac{4}{7}$.

In the following, we sketch a way to find the optimal stationary contract $\sigma(\bar{a}^e, \bar{a}^1, \bar{a}^2)$ for all discount factors. We start with the action profile that maximizes the joint payoff $G(a)$ among all $a \in A$ as our candidate for $\bar{a}^e$; here this is (L, L). This profile will either turn out to be admissible for discount factors greater than some critical value, or for no discount factor, then we would try the action profile with the next highest joint payoff. Similarly, for player i's punishment we have to start with an action profile a^i that minimizes player i's cheating payoff $c_i(a)$ over all possible action profiles $a \in A$. Since the cheating payoff only depends on the punisher's action a^i_j, we must make a choice for a^i_i. It is obvious from the constraints (5.8) and (5.9) that we have to check only the action that maximizes $G(a^i) - c_j(a^i)$. Our candidates are thus $\bar{a}^1 = (M, H)$ and $\bar{a}^2 = (H, M)$. A short calculation reveals that the action profiles $(L, L), (M, H), (H, M)$ indeed define a stationary contract for $\delta \geq \frac{1}{3}$. Side payments here facilitate collusion for intermediate discount factors because the payments can be used to smoothly adjust the punishment.

To find optimal contracts for the remaining discount factors, note first that the binding condition was the incentive constraint for the equilibrium path (condition (5.8)). Whenever the equilibrium constraint is binding, we have to replace the candidate for equilibrium play with the action profile that generates the next highest joint payoff among all profiles that relax this constraint, using the same punishment actions as before. In this symmetric game, it does not matter whether we pick (L, M) or (M, L) as our next candidate. The equilibrium condition (5.8) then holds for all $\delta \geq \frac{2}{11}$, while the punishment constraints in (5.9) hold for all $\delta \geq \frac{1}{4}$. Thus, for all $\delta \in [\frac{1}{4}, \frac{1}{3})$ at least a partially collusive outcome can be sustained. Note that a stationary contract of the form $\sigma((L, M), (M, H), (H, M))$ requires positive equilibrium payments from firm 2 to firm 1.

Since this time it is the constraint for the punishment path (5.9) that is binding, high output is not sustainable for lower discount factors. Whenever condition (5.9) is binding for player i, the punishment action profile for player i has to be replaced by the profile that offers the lowest player i cheating payoff among all profiles that relax the binding condition. In our case, the punishment actions have to be replaced by the Nash equilibrium (M, M). Since punishment with medium quantity is not sufficient for equilibrium play of (L, M), for $\delta < \frac{1}{4}$ the optimal contract is repeated play of (M, M), which is the action profile with the next highest joint payoff.

5.4 Renegotiation-proofness: Strong perfection

So far we have characterized Pareto optimal subgame perfect outcomes, which seem likely outcomes if the players can communicate and thus would not some-how be trapped in an inefficient equilibrium. However, if the players can meet and coordinate on an efficient subgame perfect equilibrium at the beginning of their relationship, they may be able to do the same after any history for which the original continuation equilibrium does not lead to a Pareto optimal outcome. If the players anticipate such a renegotiation, the equilibrium strategies may cease to be credible. Therefore, one may predict that players will initially choose an equilibrium that is immune to this criticism. For example, there is never scope for renegotiation if an equilibrium creates the same maximum surplus $\bar{U}_{SGP}$ after every possible history.

This strongest notion of renegotiation proofness is called strong optimality in Levin (2003). Levin introduces this criterion in his analysis of Pareto optimal relational contracts in a principal-agent setting where it is easily fulfilled. In fact, he shows that in his game all Pareto optimal subgame perfect payoffs can be implemented by a strongly optimal equilibrium. This result depends, however, on the way Levin defines this notion for a repeated sequential game. He assumes that negotiations only take place at the beginning of a period, but never just before a punishment has to be carried out.

Similarly, if in our setting we define a subgame perfect equilibrium σ to be strongly optimal if $U(\sigma|h) = \bar{U}_{SGP}$ for every history $h \in H^{pay}$, then all optimal stationary contracts $\sigma(\bar{a}^e, a^1, a^2)$ are strongly optimal. The reason is that in every continuation equilibrium starting with the side payment, the required payments will be conducted and equilibrium play $\bar{a}^e$ is continued or resumed. Since by assumption there is no renegotiation directly before the play stage, continuation equilibria that require inefficient play are never subject to renegotiation. This implies that the set of strongly optimal payoffs coincides with the Pareto frontier of subgame perfect payoffs $\mathcal{P}(\mathcal{U}_{SGP}^{pay})$.

The assumption that no renegotiation is possible before the action stage is natural if payments do not have to be carried out at a certain point in time, such that not paying continues to be a reversible action. In this case, the threat that implements a compensation scheme can never be negotiated away from. Any try to renegotiate the punishment will be blocked by the nondeviating party, who believes that it will receive the fine as soon as the negotiations stop. The focal disagreement point is always the original equilibrium, and therefore the deviating party cannot credibly argue that it has no intention of paying.

For a more stringent test of renegotiation proofness one may want to analyze a different timing, and admit renegotiation between the two stages of a period, which we will do in the remainder of this chapter. With this timing there can only in special cases be an equilibrium that creates joint surplus $\bar{U}_{SGP}$ after every history $h \in H$, since punishments typically entail some efficiency loss. We therefore turn to the slightly weaker concept of strong perfection, which requires that no continuation payoff is strictly Pareto dominated by another subgame perfect payoff.[8]

Definition 5.5. *A subgame perfect equilibrium* σ *is* strong perfect *also within periods if* $\mathcal{U}^k(\sigma) \subset \mathcal{P}(\mathcal{U}^k_{SGP})$ *for all* $k \in \{pay, play\}$.

Strong perfect equilibria may fail to exist, but the concept provides a useful sufficient condition for renegotiation-proofness. If every other subgame perfect equilibrium makes at least one of the players worse off, then one may feel confident that renegotiation is deterred. We show next that every strong perfect payoff can be sustained by a stationary contract, and then use this fact to characterize the set of strong perfect payoffs.

Proposition 5.6. *If the set of strong perfect payoffs* $\mathcal{U}^{pay}_{SP}$ *is nonempty, then there exists a strong perfect optimal stationary contract* $\sigma(\bar{a}^e, a^1, a^2)$ *with punishment payoffs* u^i_i *such that* $\mathcal{U}^{pay}_{SP}$ *is the line between* $(u^1_1, G(\bar{a}^e) - u^1_1)$ *and* $(G(\bar{a}^e) - u^2_2, u^2_2)$.

Proof. See Section 5.7 at the end of the chapter.

An optimal stationary contract can only in the punishment phases, before the play stage, be dominated by another subgame perfect equilibrium. To find a characterization of strong perfect stationary contracts, we are therefore interested in $\mathcal{P}(\mathcal{U}^{play}_{SGP})$. Let $\sigma \in \Sigma^{play}_{SGP}$ with $u(\sigma) \in \mathcal{P}(\mathcal{U}^{play}_{SGP})$ be a Pareto optimal equilibrium in the subgame that starts with the play stage. Let $\tilde{a}$ be the first action profile on the equilibrium path of σ. Clearly, there exists an optimal stationary contract $\sigma(\bar{a}^e, \bar{a}^1, \bar{a}^2)$ that weakly dominates $u(\sigma|\tilde{a})$. It can then be shown that play of $\tilde{a}$ followed by play of $\sigma(\bar{a}^e, \bar{a}^1, \bar{a}^2)$ must be a subgame perfect equilibrium with payoffs equal to $u(\sigma)$. Equilibria of this form are essential for the shape of $\mathcal{P}(\mathcal{U}^{play}_{SGP})$. We define:

Definition 5.7. *An action profile* $\tilde{a} \in A$ *is called* admissible *if*

$$(1-\delta)G(\tilde{a}) + \delta \bar{U}_{SGP} \geq (1-\delta)(c_1(\tilde{a}) + c_2(\tilde{a})) + \delta(\bar{u}^1_{SGP} + \bar{u}^2_{SGP}).$$

The preceeding discussion suggests that an action profile $\tilde{a}$ is admissible if and only if there is a subgame perfect equilibrium in which $\tilde{a}$ is played on the equilibrium path. Admissible actions play an important role in the following characterization of strong perfect stationary contracts.

---

[8] That we use strict Pareto dominance matters for knife-edge cases only, for which the conclusion of renegotiation proofness then is not as strong.

Proposition 5.8. *An optimal stationary contract $\sigma(\bar{a}^e, a^1, a^2)$ with punishment payoffs u_1^1 and u_2^2 is strong perfect if and only if for both players $i \in \{1, 2\}$ and all admissible $\tilde{a}$ with $G(\tilde{a}) > G(a^i)$ it holds that either*

(i) $(1-\delta)\left(G(\tilde{a}) - c_i(\tilde{a})\right) - \delta \bar{u}^i_{SGP} \leq (1-\delta)G(a^i) - u_i^i$, *or*
(ii) $(1-\delta)(G(\tilde{a}) - c_j(\tilde{a})) + \delta G(\bar{a}^e) - \delta \bar{u}^j_{SGP} \leq u_i^i$ *for* $j \neq i$.

Proof. See Section 5.7 at the end of the chapter.

Intuitively, conditions (i) and (ii) concern the punishment for player i in a stationary contract: condition (i) ensures that there are no subgame perfect equilibria with higher payoff for the other player, and if this condition fails, condition (ii) adds that such equilibria should not have a higher player i payoff. [9] We now derive corollaries that are easier to apply than the proposition:

Corollary 5.9. *There exists a strong perfect stationary contract with punishment payoffs $\bar{u}^i_{SGP}$ - and consequently $\mathcal{U}^{pay}_{SP} = \mathcal{P}(\mathcal{U}^{pay}_{SGP})$ - if and only if for $i = 1, 2$ there exists an admissible $\bar{a}^i \in A$ with $c_i(\bar{a}^i) = \bar{u}^i_{SGP}$ such that*

$$G(\bar{a}^i) - c_i(\bar{a}^i) \geq G(\tilde{a}) - c_i(\tilde{a})$$

for all admissible $\tilde{a} \in A$.

Proof. See Section 5.7 at the end of the chapter.

Corollary 5.10. *There exists no strong perfect stationary contract $\sigma(\bar{a}^e, a^1, a^2)$ if for both players $i = 1, 2$ it holds that $G(\bar{a}^e) > G(a^i)$ and*

$$(1-\delta)\left(G(\bar{a}^e) - c_i(\bar{a}^e)\right) - \delta \bar{u}^i_{SGP} > (1-\delta)G(a^i) - c_i(a^i).$$

Proof. See Section 5.7 at the end of the chapter.

These two corollaries can be used to show that in Abreu's Cournot example there is no strong perfect equilibrium, except for the Nash equilibrium of the stage game as the only admissible action profile in case $\delta < \frac{1}{4}$. Instead of exercising this non-existence in detail, we now present examples in which strong perfect equilibria at least sometimes exist.

---

[9] That the latter case cannot be excluded in a strong perfect stationary contract, and that thus we cannot restrict attention to strong perfect stationary contracts with maximum fines, is due to the fact that we did not assume the existence of a public correlation device. In a stationary contract that ceases to be strong perfect if the maximum fines are used, the punishment payoff is always dominated by a convex combination of payoffs in $\mathcal{U}^{play}_{SGP}$.

Example: Principal-Agent Game

Assume that only player 1 (the agent) chooses an action $a \in A_1$. The action creates a nonpositive payoff $g_1(a)$ for player 1 and a nonnegative benefit $g_2(a)$ for player 2 (the principal). One interpretation is that player 1 is a supplier who delivers a product of a certain quality, where higher quality is more expensive. Another interpretation is that player 1 is a worker who can exert work effort a, which can be observed by the employer. The agent can choose a 'do-nothing' action $a = 0$ that yields zero payoff for both players.

Since $a = 0$ is the worst punishment for player 2, while any action yields zero cheating payoff to player 1, the subgame perfection condition is $G(a) \geq (1-\delta)g_2(a)$. The optimal equilibrium action $\bar{a}^e$ is therefore the one that maximizes $G(a)$ among all $a \in A_1$ with $G(a) \geq (1-\delta)g_2(a)$. We use corollary 5.9 to show that $\sigma(\bar{a}^e, \bar{a}^e, 0)$ is strong perfect. The punishment for player 1 is simply that he does not get paid, while the punishment for player 2 if he does not pay is termination of the relationship for the current period. Since for any admissible action $\tilde{a}$ it holds that $G(\tilde{a}) - g_2(\tilde{a}) = g_1(\tilde{a}) \leq 0$, the equilibrium is strong perfect. Hence, in this simple complete information game, we confirm the intuition of Levin (2003) that with a one-sided incentive problem, optimal subgame perfect payoffs can be implemented in a renegotiation-proof way.[10]

Example: Strong Perfection in the Prisoner's Dilemma

For another example, consider a Prisoner's Dilemma game of the form

	C	D
C	$1,1$	$d, S-d$
D	$S-d, d$	$0,0$

with $d > 1 > \frac{S}{2}$. Cooperation can be sustained if $\delta \geq \frac{d-1}{d} =: \delta_{CC}$.

First, assume that $S \leq 0$. In this case, (D, D) is not only an easier punishment, but also the more efficient one. For $\delta < \delta_{CC}$ only the equilibrium (D, D) is subgame perfect, and therefore of course also renegotiation-proof. For $\delta \geq \delta_{CC}$, we use Corollary 5.9 to see that there is a strong perfect stationary contract with equilibrium action (C, C) if and only if $d \geq 2$. If this condition holds, $\mathcal{U}_{SP}^{pay} = \{(u_1, 2-u_1) : u_1 \in [0, 2]\}$. It may seem counterintuitive that for strong perfection the temptation to deviate from (C, C) in the stage game must be sufficiently large, but this condition implies that payoffs on the equilibrium path cannot be too asymmetric. At the end of section 5.5

[10] That punishment in which the agent is punished by not being paid, while continueing to work, can lead to strong perfect equilibria has already been noted by Farrell and Weizsäcker (2001).

we will return to the prisoners' dilemma and compare this condition for strong perfection to the condition for strong renegotiation proofness.

Next, take the case that $S > 0$ and $d \geq 2 - S$. Note that for all $\delta \geq \delta_{CC}$, (C, D) and (D, C) are also admissible, and corollary 5.9 again tells us that the resulting $\sigma((C, C), (C, D), (D, C))$ are strong perfect. For $\delta < \delta_{CC}$ we have to distinguish two cases. If it happens to be true that $S \leq 1$, then only (D, D) is admissible for these low discount factors. If instead $S > 1$ then there is a range of discount factors for which (C, D) is admissible but (C, C) is not. In this case, stationary contracts of the form $\sigma((C, D), (C, D), (D, C))$ are strong perfect.

Finally, for the case that $S > 0$ and $2 > d + S$ there are only trivial strong perfect equilibria: (D, D) as the only stationary contract for any $\delta < \delta_{CC}$.

5.5 Renegotiation proofness: WRP and SRP

Strong perfection is a very strong criterion; in a strong perfect equilibrium every continuation payoff must survive comparison to all subgame perfect equilibria, even to those that are not considered possible ways of playing the game themselves. In this section, we analyze a concept that tries to avoid such comparisons, namely weak and strong renegotiation-proofness defined by Farell and Maskin (1989). An equilibrium is weakly renegotiation-proof if none of its continuation equilibria is strictly Pareto dominated by another continuation equilibrium. Strong renegotiation proofness requires stability against renegotiation to any possible weakly renegotiation proof continuation equilibrium. The formal definitions, allowing for renegotiation within a period, are as follows:

Definition 5.11. *A subgame perfect equilibrium σ is weakly renegotiation proof (WRP), if for no stage $k \in \{pay, play\}$ there are continuation payoffs $u, u' \in \mathcal{U}^k(\sigma)$ such that u is strictly Pareto dominated by u'.*

WRP equilibria always exist in our framework, but the concept often does not have much restricting power. For example, playing the same Nash equilibrium of the stage game after every history is always a WRP equilibrium. If an equilibrium σ is WRP, then we also say that the payoffs $\mathcal{U}^k(\sigma)$ are WRP. Let $\mathcal{U}^k_{WRP}$ denote the set of all WRP payoffs beginning with stage k.

Definition 5.12. *A WRP equilibrium σ is strongly renegotiation proof (SRP) if for no stage k and $u \in \mathcal{U}^k(\sigma)$ there exists another WRP payoff $u' \in \mathcal{U}^k_{WRP}$ such that u is strictly Pareto dominated by u'.*

It follows directly from this definition that the set of SRP payoffs is a subset of the Pareto-frontier of all WRP payoffs, but in general the two sets do not coincide. In fact, SRP equilibria often do not even exist. When monetary transfers are possible, however, Baliga and Evans (2000) show in a related model that as players grow infinitely patient, the set of SRP payoffs approaches the set of all efficient total payoff in the stage game payoffs, which are the individually rational distributions of the maximum surplus that can be achieved in the stage game. This insight carries through to our framework. In the following we characterize the set of SRP payoffs for all discount factors $\delta \geq \frac{1}{2}$. This assumption that the future matters more than the present guarantees that the way we specified the payments in a stationary contract is general enough to characterize SRP payoffs. At the end of this section, we provide conditions for the results to extend to the case $\delta < \frac{1}{2}$.

First, we answer the question under what conditions a given stationary contract $\sigma(a^e, a^1, a^2)$ is WRP. We do this only for a subclass of stationary contracts that is sufficient for our analysis.

Definition 5.13. *A stationary contract $\sigma(a^e, a^1, a^2)$ is called* regular *if it has maximum fines, $G(a^e) \geq G(a^i)$, and either $c_i(a^i) < c_i(a^e)$ or $a^i = a^e$ for $i = 1, 2$.*

Lemma 5.14. *Let $\sigma(a^e, a^1, a^2)$ with equilibrium payoffs $u^e = g(a^e) - \delta p^e$ be a regular stationary contract. Then $\sigma(a^e, a^1, a^2)$ is WRP if and only if*

$$(1-\delta)G(a^i) + \delta G(a^e) - c_i(a^i) \geq u_j^e \text{ for } i \neq j = 1, 2. \tag{5.10}$$

There exists a payment vector $\tilde{p}^e$ such that $\sigma(a^e, a^1, a^2)$ with equilibrium payments $\tilde{p}^e$ instead of p^e is WRP if and only if

$$(1-\delta)G(a^i) + \delta G(a^e) - c_i(a^i) \geq (1-\delta)c_j(a^e) + \delta c_j(a^j) \tag{5.11}$$

$$(1-\delta)(G(a^1) + G(a^2)) + 2\delta G(a^e) \geq G(a^e) + c_1(a^1) + c_2(a^2). \tag{5.12}$$

The intuition is simple. Since in a stationary contract all continuation equilibria starting with the side payment stage have the same total payoff, they cannot dominate each other. In the regular case in which the largest surplus is generated on the equilibrium path, the only inefficiency occurs when the actual punishment has to be carried out. If a player's punishment is his least preferred outcome, for WRP each player has to prefer the other's punishment to the equilibrium path. The conditions (5.11) and (5.12) ensure that it is possible to choose the equilibrium transfer such that this holds.

If the game is symmetric and in addition $a^e = (a_1^e, a_1^e)$, side payments on the equilibrium path are superfluous; in this case one can show that actions

$a^e, a^1, a^2 = (a_2^1, a_1^1)$ can be implemented in a regular stationary WRP contract if and only if it holds that

$$(1-\delta)G(a^1) - c_1(a^1) \geq (1-2\delta)g_1(a^e). \tag{5.13}$$

Simple conditions for stationary contracts to be WRP are very useful, because they allow us to characterize the Pareto frontier of WRP payoffs.

Proposition 5.15. *Let $\delta \geq \frac{1}{2}$. For all $\sigma \in \Sigma_{WRP}^{pay}$ there exists a regular WRP stationary contract $\sigma(a^e, a^1, a^2)$ with $G(a^e) \geq u_1 + u_2$, $c_i(a^i) \leq u_i$ and $G(a^i)(1-\delta) + \delta G(a^e) - c_i(a^i) \geq u_j$ for all $u \in \mathcal{U}^{play}(\sigma)$.*

Proof. See Section 5.7 at the end of the chapter.

In particular, we may restrict our attention to regular WRP stationary contracts when we are interested in the WRP payoff that maximizes the joint surplus of the players. We say that a regular WRP stationary contract $\sigma(a^e, a^1, a^2)$ is an optimal WRP contract if $G(a^e) \geq U(\sigma)$ for all $\sigma \in \Sigma_{WRP}^{pay}$. Moreover, it follows from the proposition that any payoff on the Pareto frontier $\mathcal{P}(\mathcal{U}_{WRP}^{pay})$ can be implemented by a regular WRP stationary contract.

While in a WRP equilibrium the punishment need not fit the crime, it must fit the equilibrium play. Hence, there does not have to be an equilibrium which can be used as punishment in all WRP equilibria. This changes when we turn to SRP equilibria, in which all continuation payoffs must lie on the Pareto frontier of WRP payoffs. Since elements of $\mathcal{P}(\mathcal{U}_{WRP}^{play})$ do not Pareto dominate each other, there is now a chance to find universal punishments.

Proposition 5.16. *Let $\delta \geq \frac{1}{2}$. If the set of SRP payoffs is nonempty, there exists an SRP stationary contract $\sigma(a^e, a^1, a^2)$ with punishment payoffs u_1^1 and u_2^2 such that $\mathcal{U}_{SRP}^{pay}$ is given by the line between $(u_1^1, G(a^e) - u_1^1)$ and $(u_2^2, G(a^e) - u_2^2)$.*

Proof. See Section 5.7 at the end of the chapter.

This proposition is only helpful if we know whether SRP equilibria exist at all. Fortunately, we can prove the following sufficient condition for an optimal WRP stationary contract to be SRP:

Proposition 5.17. *Let $\delta \geq \frac{1}{2}$. A regular optimal WRP contract $\sigma(a^e, a^1, a^2)$ is SRP if there is no regular WRP stationary contract $\sigma(\tilde{a}^e, \tilde{a}^1, \tilde{a}^2)$ and $i \in \{1, 2\}$ such that*

$$G(\tilde{a}^i)(1-\delta) + \delta G(\tilde{a}^e) - c_i(\tilde{a}^i) > G(a^i)(1-\delta) + \delta G(a^e) - c_i(a^i).$$

Proof. See Section 5.7 at the end of the chapter.

Remarks on the case $\delta < \frac{1}{2}$

For many games, our characterization of WRP and SRP payoffs extends beyond the case $\delta \geq \frac{1}{2}$, because the following holds:

Remark 5.18. If the actions of every regular stationary contract $\sigma(a^e, a^1, a^2)$ that satisfy conditions (5.11) also satisfy (5.12), then the results in Propositions 5.15, 5.16, and 5.17 also hold for the case $\delta < \frac{1}{2}$.

This follows immediately because Propositions 5.16 and 5.17 assume that $\delta \geq \frac{1}{2}$ only because they rely on Proposition 5.15. That Proposition 5.15 holds also for $\delta < \frac{1}{2}$ if the joint WRP condition (5.12) follows from the other conditions is obvious from its proof.

In the following we sketch how to extend the definition of a stationary contract in order to be able to implement all WRP payoffs on the Pareto frontier. So far, we have assumed that in a stationary contract all transfers add up to zero. Now we also admit the possibility that for the equilibrium transfer p^e it holds that $p_1^e + p_2^e < 0$. We call a stationary contract with $p_1^e + p_2^e < 0$ a money-burning stationary contract. Such an artificial inefficiency may be necessary to find all WRP payoffs, because play on the punishment path is compared to play on the equilibrium path. If it is not possible to make the punishment more efficient, instead resources have to be destroyed on the equilibrium path.

The conditions for weak renegotiation-proofness of a stationary contract with money burning are more involved than those for a regular stationary contract. Let $\delta < \frac{1}{2}$ and $\sigma(a^e, a^1, a^2)$ be a stationary contract such that the actions a^e, a^1, a^2 fulfill $G(a^e) \geq G(a^i)$, $c_i(a^i) < c_i(a^e)$ as well as the conditions in (5.11), but not condition (5.12). The highest possible total payoff in a money-burning WRP $\sigma(a^e, a^1, a^2)$ then is

$$\bar{U} = \frac{(1-\delta)(G(a^1) + G(a^2)) - (c_1(a^1) + c_2(a^2))}{1 - 2\delta} \leq G(a^e). \tag{5.14}$$

We can define the transfer p^e such that $u_i^e := G(a^j)(1-\delta) + \delta\bar{U} - c_j(a^j)$. Note that $u_1^e + u_2^e = \bar{U}$.

We have to make sure that the reduced joint payoff on the equilibrium path still satisfies the subgame perfection conditions, which take the form

$$(1-\delta)G(a^i) + \delta\bar{U} - c_i(a^i) \geq (1-\delta)\max(c_j(a^i), c_j(a^e)) + \delta c_j(a^j). \tag{5.15}$$

The WRP conditions for payoffs in the play stage hold by definition, but we also have to take into account that continuation payoffs before the payment stage now do not all have the same sum of payoffs. In a continuation equilibrium before the fine or the adjustment payment is paid the total payoff is

$\bar{U}$, while it is lower before the equilibrium transfer. Therefore, we only have to ensure that the continuation payoff before the equilibrium transfer is not dominated by the continuation payoff before the adjustment payment is made, i.e., that either

$$u_i^e - (1-\delta)g_i(a^e) \geq c_i(a^i) - (1-\delta)g_i(a^i)$$

or

$$u_j^e - (1-\delta)g_j(a^e) \geq \delta\bar{U} - c_i(a^i) + (1-\delta)g_i(a^i),$$

which is equivalent to

$$\bar{U} \geq g_j(a^i) + g_i(a^e) \text{ or } g_j(a^i) \geq g_j(a^e) \tag{5.16}$$

for $i \neq j \in \{1, 2\}$.

Proposition 5.19. *For all $u \in \mathcal{U}_{WRP}^{pay}$ there exists a WRP, possibly money-burning, stationary contract with maximum fines that weakly Pareto dominates u.*

Proof. See Section 5.7 at the end of the chapter.

All payoffs on the strict Pareto frontier of WRP payoffs can be implemented by a WRP stationary contract with money burning. Unfortunately, for the case that $\delta < \frac{1}{2}$, $\mathcal{P}(\mathcal{U}_{WRP}^{pay})$ does not have to be linear. Characterization of SRP payoffs is therefore difficult if money burning stationary contracts are needed, and since these cases are more the exception than the rule, we do not attempt to go into more detail here. Note, however, that there is no a priori reason why such a money burning stationary contract cannot be SRP. Since continuation equilibria need only be compared to other WRP payoffs, they never have to be compared to the relational contract that just skips the money burning and is otherwise identical, because this one is not WRP.

While money burning can never be part of a strong perfect relational contract, it arises naturally when requiring weak renegotiation-proofness. This is in part due to the timing of renegotiations; if it were possible to separate the money burning from the transfer, and renegotiate from before the money burning to the point where the money has already been burned, then such money burning stationary contracts would also not be WRP. Nevertheless, this type of stationary contract does fit in with the spirit of weak renegotiation-proofness, which requires that equilibrium outcomes may not be too desirable compared to the payoffs in the punishments. Moreover, money buring is used here only as a means to find WRP payoffs, and as such does not have to resemble real-world agreements. For every WRP payoff there may be many different equilibria that implement this payoff, possibly including some that do not require money burning on the equilibrium path.

Examples

First, we go back to the examples of Section 5.3 and 5.4, to find that the SRP criterion excludes equilibria with punishment actions that create a low joint surplus. In Abreu's Cournot game, we have that for $\delta \geq \frac{1}{3}$, the optimal stationary contract $\sigma((L, L), (L, H), (H, L))$ is SRP, while for $\delta \in [\frac{4}{13}, \frac{1}{3})$, all contracts of the form $\sigma((L, M), (L, H), (H, L))$ are SRP. In contrast, stationary contracts with punishments (M, H) or (H, M) are never SRP. For $\delta < \frac{4}{13}$ only the Nash equilibrium (M, M) is WRP, and hence SRP.

In the Prisoner's dilemma, for $\delta \geq \frac{1}{2}$ every optimal stationary contract that uses the punishment with the highest joint payoff is SRP. To understand why this condition is so different from the condition for strong perfection, consider again the case $S \leq 0$ and $\delta \geq \delta_{CC}$ to compare the condition for weak renegotiation proofness, $\delta \geq \frac{1}{2}$, to the condition for strong perfection, which was $2 \leq d$. For weak renegotiation-proofness, player $j's$ payoff in $i's$ punishment only has to be higher than continuation payoffs in the same equilibrium. Therefore the payoff 2δ has to be greater or equal to the equilibrium payoff 1, which is true if $\delta \geq \frac{1}{2}$. In contrast, for strong perfection player j's payoff has to be compared to all possible subgame perfect payoffs, including those with positive equilibrium transfers. The higher the defection payoff d is, the more of those asymmetric payoffs there are. The largest player j payoff is $2 - (1 - \delta)d$, and comparing this to 2δ we arrive at the condition $2 \leq d$. Since these asymmetric equilibria are not necessarily WRP, they do not have to be taken into account for SRP, which provides an intuition for why the SRP condition is also equal to $\delta \geq \frac{1}{2}$ for the case $S \leq 0$.[11]

The only case that is missing in the case distinctions regarding the prisoner's dilemma is $S > 0$, $2 - d > S$, $\frac{1}{2} > \delta \geq \delta_{CC}$. Let δ_{CD} denote the discount factor at which (C, D) becomes admissible. For all $\delta < \delta_{CD}$, the equilibrium (D, D) is SRP. For all $\delta \geq \frac{1-S}{2-S}$, $\sigma((C, C)(C, D), (D, C))$ is SRP, and for $\delta \in [\delta_{CD}, \frac{1-S}{2-S})$ all conditions except the joint WRP condition are satisfied. In this case, money burning is needed to find all WRP payoffs.

As an example for a money-burning stationary contract, let the variables take the values $S = \frac{1}{2}$, $d = \frac{5}{4}$, $\delta = \frac{3}{10}$. The maximal joint payoff is $\bar{U} = \frac{7}{4}$, which can be achieved by the money-burning stationary contract that has the following continuation payoffs in the play stage: $(\frac{7}{8}, \frac{7}{8})$, $(\frac{7}{8}, 0)$ and $(0, \frac{7}{8})$. Continuation payoffs before the payment stage are $(\frac{7}{12}, \frac{7}{12})$, $(\frac{7}{4}, 0)$, $(0, \frac{7}{4})$.

---

[11] The maximal payoff in any WRP equilibrium, given by 5.14, is equal to 0 for $\delta < \frac{1}{2}$.

Collusion in a Bertrand game

Bertrand competition with symmetric costs

We now investigate the case of a Bertrand duopoly with side payments. In order to have a compact strategy space and well-defined cheating payoffs, we assume that prices are chosen from a finite grid $A_i = \{\epsilon m\}_{m=0}^{\bar{m}}$, with ϵ small and $\bar{m}$ sufficiently large. Firm i's profits are given by

$$g_i(a) = \begin{cases} (a_i - k_i)D(a_i) & \text{if } a_i < a_j \\ (a_i - k_i)D(a_i)/2 & \text{if } a_i = a_j \\ 0 & \text{if } a_i > a_j \end{cases}$$

where D is a weakly decreasing market demand function, and $k_i \geq 0$ is the constant marginal cost of firm i. We first consider the case of symmetric firms $k_1 = k_2$. Clearly, marginal cost pricing is an optimal punishment, and for $\delta \geq \frac{1}{2}$ the Pareto frontier of subgame perfect payoffs includes all distributions of the monopoly profit. Using Lemma 5.14 and Prop. 5.16 one can establish that these payoffs are also SRP.

In contrast, for a Betrand duopoly without side payments, Farell and Maskin (1989) show that only marginal cost pricing can be sustained in a WRP equilibrium in pure strategies. Based on this result, McCutcheon (1997) argues that small fines for meetings of competitors can facilitate collusion since renegotiation becomes harder. However, weak renegotiation proofness only restricts the set of outcomes if mixed strategies are not allowed (Farell and Maskin, 1989) and if prices are not restricted to lie on a sufficiently coarse grid (Andersson and Wengström, 2007). Our example shows that in addition it needs to be assumed that side payments are impossible.

Bertrand competition with asymmetric costs

Miklos-Thal **?** shows that cost asymmetries facilitate collusive subgame perfect equilibria in a repeated Bertrand competition if side payments are possible. In the following, we use our characterizations to first replicate this result for two firms and then show that weak renegotiation proofness does not restrict the set of equilibrium payoffs. Let now $k_1 \leq k_2$ and $\pi_i(p) = (a_i - p)D(p)$ denote firm i's profit if it serves the whole market at a price p. As punishment profiles we choose $a^i = (k_i, k_i + \epsilon)$. Collusion is easiest to sustain if the low cost firm supplies the whole market and compensates the high cost firm with side payments. Define $\phi(p) = \frac{p-k_2}{p-k_1}$ as the ratio of firm 2's markup to firm 1's markup at price p. The condition for action profiles of the form $a^e = (a_1^e, a_1^e + \epsilon)$ to be part of a stationary contract $\sigma(a^e, a^1, a^2)$ is equal to

$$\delta \geq \frac{\phi(a_1^e)}{1 + \phi(a_1^e)} \tag{5.17}$$

as $\epsilon \to 0$. Since $\phi(a_1^e) \leq 1$, cost asymmetries facilitate collusion. Moreover, since $\phi(k_2) = 0$, some collusive markup can be sustained for every discount factor $\delta > 0$. Such contracts $\sigma(a^e, a^1, a^2)$ are WRP if

$$\delta \geq \frac{\pi_1(a_1^e) - \pi_1(k_2)}{2\pi_1(a_1^e) - \pi_1(k_2)}. \tag{5.18}$$

For a perfectly inelastic demand function D, condition (5.18) is identical to the subgame perfection condition (5.17), and since D is weakly decreasing, the first condition always implies the latter.

5.6 Concluding remarks

We have shown that Pareto optimal subgame perfect payoffs and renegotiation-proof payoffs can generally be found by restricting attention to a class of stationary contracts. These stationary contracts prescribe play of the same action in every period on the equilibrium path, and in case of a deviation allow the deviator to pay a fine and return to equilibrium play. The actual punishment that results if a fine is not paid occurs within one period. Fines can hence be used to create one-period punishments, which are more resistent to renegotiation than a punishment that lasts forever. Renegotiations are often blocked by the expectations for the next period.

In fact, a one period inefficiency will also never be renegotiated if negotiations only take place after actions have already been taken, i.e., if the timing follows a "bargain-pay-play" pattern. This timing can be thought of as bargaining and action occuring at certain points in time while payments can in principle be made all the time. In contrast, the "bargain-pay-bargain-play" timing that is mainly analyzed the present chapter implies that payment and action occur at specified dates, but bargaining can occur at all times. Since in applications it is often difficult to pin down the timing of negotiations, the most important results in this paper are probably the sufficient conditions for strong perfection, because this concept offers a similarly strong conclusion of renegotiation-proofness as Levin (2003)'s strong optimality. These results can therefore serve the purpose of a robustness-against-renegotiation-check for optimal subgame perfect equilibria.

The results in this chapter add to the discussion in Fong and Surti (2009) about the role of side payments in relational contracts. Among other things, they conjecture that restriction to stationary equilibrium paths without correlation is without loss of generality also beyond the prisoners' dilemma that they study. We have shown here that even with possibly complex punishments, stationary contracts are sufficient to characterize Pareto optimal subgame perfect payoffs.[12] Fong and Surti (2009) do not attempt to make a conjecture about the generalizability of their results on renegotiation proofness, which are complex and technical in their framework with an impatient and a patient player. In contrast, in our framework with a common discount factor, finding the set of strong perfect, WRP, or SRP payoffs is often not difficult, and the use of stationary contracts helps to explain the technical conditions that describe these sets. To see whether and how these characterizations extend to more general frameworks, e.g., with private information or costly transfers, are interesting tasks for future work.

[12] The proof can be extended to show that correlation devices are not needed for this task if side payments, including an ex ante payment, are possible. The existence of such devices may, however, matter for the set of renegotiation-proof outcomes. We have excluded the use of a public correlation device in the analysis of renegotiation-proof payoffs because it is not clear what this would mean for the timing of negotiations.

5.7 Proofs

Proof of Lemma 5.3.
We are interested in finding conditions on a^e, a^1, a^2 that make it possible to define the equilibrium transfer p^e and fines F^1 and F^2 such that conditions (5.4), (5.5), and (5.6) for subgame perfection are fulfilled. Note that there are three conditions that bound u_i^i, $i = 1, 2$ from above but only one that bounds it from below. Therefore, these conditions hold for some u_i^i if and only if they hold for the lowest continuation payoff $u_i^i = c_i(a^i)$. Since the equilibrium transfer p^e only plays a role in the equilibrium conditions

$$g_i(a^e) - \delta p_i^e \geq c_i(a^e)(1-\delta) + \delta c_i(a^i) \ \text{ for } i \in \{1, 2\},$$

we can for example choose $\delta p_1^e = g_1(a^e) - c_1(a^e)(1-\delta) - \delta c_1(a^1)$ such that player 1's incentive constraint is binding. Player 2's incentive constraint then reads

$$G(a^e) \geq (c_1(a^e) + c_2(a^e))(1-\delta) + \delta(c_1(a^1) + c_2(a^2)),$$

which is just the sum of the two equilibrium incentive constraints and must therefore also hold if those hold separately.

Proof of Proposition 5.4.
There must be sequences of equilibria in Σ_{SGP}^{play}, $\left(\sigma_n^e, \sigma_n^1, \sigma_n^2\right)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} U(\sigma_n^e) = \bar{U}_{SGP}$ and $\lim_{n\to\infty} u_i(\sigma_n^i) = \bar{u}_{SGP}^i$. Let a_n^i be the first action profile of the equilibrium path of σ_n^l, $l = e, 1, 2$. Then a_n^l is a sequence in the compact set $A = A_1 \times A_2$, and as such must have convergent subsequences with limits in A. We assume here w.l.o.g. that these subsequences are already given by a_n^l and denote the limits by $\bar{a}^e, \bar{a}^1$ and $\bar{a}^2$, resp. In the following we use the properties of $\sigma_n^e, \sigma_n^1, \sigma_n^2$ to make inferences about the limit actions. First, if we decompose $U(\sigma_n^e)$ into current period payoff and future payoff, we find that

$$U(\sigma_n^e) \leq G(\bar{a}_n^e)(1-\delta) + \delta \bar{U}_{SGP}.$$

Since G is continuous, taking limits on both sides yields

$$\bar{U}_{SGP} \leq G(\bar{a}^e). \tag{5.19}$$

Second, subgame perfection of σ_n^i implies

$$u_i(\sigma_n^i) \geq c_i(a_n^i)(1-\delta) + \delta \bar{u}_{SGP}^i.$$

Again, since c_i is continuous, taking limits yields

$$\bar{u}_{SGP}^i \geq c_i(\bar{a}^i). \tag{5.20}$$

Third, summing up player 1 and 2 's subgame perfection conditions for σ_n^e yields

$$\bar{U}_{SGP} \geq (c_1(a_n^e) + c_2(a_n^e))\,(1-\delta) + \delta(\bar{u}^1_{SGP} + \bar{u}^2_{SGP}).$$

In the limit, and using (5.19) and (5.20), this becomes

$$G(\bar{a}^e) \geq (c_1(a^e) + c_2(a^e))\,(1-\delta) + \delta(c_1(\bar{a}^1) + c_2(\bar{a}^2)). \tag{5.21}$$

Last, we exploit the subgame perfection condition

$$u_j(\sigma_n^i) \geq c_j(a_n^i)(1-\delta) + \delta \bar{u}^j_{SGP}$$

as well as

$$G(a_n^i)(1-\delta) + \delta \bar{U}_{SGP} \geq U(\sigma_n^i)$$

to get

$$G(a_n^i)(1-\delta) + \delta \bar{U}_{SGP} - \bar{u}^i_{SGP} \geq c_j(a_n^i)(1-\delta) + \delta \bar{u}^j_{SGP}.$$

In the limit, and using (5.19) and (5.20), this becomes

$$G(\bar{a}^i)(1-\delta) + \delta G(a^e) - c_i(\bar{a}^i) \geq c_j(\bar{a}^i)(1-\delta) + \delta c_j(\bar{a}^j). \tag{5.22}$$

Equations (5.21) and (5.22) together with Lemma 5.3 now tell us that there is a stationary contract $\sigma(\bar{a}^e, a^1, a^2)$, with joint payoff $G(\bar{a}^e) = \bar{U}_{SGP}$ and punishment payoffs $c_i(a^i) = \bar{u}^i_{SGP}$. The up-front payment can be used to achieve all payoffs on the line between $(c_1(\bar{a}^1), G(\bar{a}^e) - c_1(\bar{a}^1))$ and $(G(\bar{a}^e) - c_2(\bar{a}^2), c_2(\bar{a}^2))$.

Proof of Proposition 5.6.
Assume that $\mathcal{U}_{SP}^{play}$ is nonempty. Let $\bar{u}^i$ be a tuple in the closure of $\mathcal{U}_{SP}^{play}$ with $\bar{u}^i_i = \inf_{u \in \mathcal{U}_{SP}^{play}} u_i$, $i = 1, 2$. Since the lowest possible strong perfect payoffs must be able to implement at least one action profile $\bar{a}^e$ with $G(\bar{a}^e) = \bar{U}_{SGP}$ it must hold that

$$G(\bar{a}^e) \geq (c_1(\bar{a}^e) + c_2(\bar{a}^e))\,(1-\delta) + \delta(\bar{u}^1_1 + \bar{u}^2_2).$$

It also holds again that $\bar{u}^i_i = \inf_{u \in \mathcal{U}_{SP}^{pay}} u_i$.

As in the SGP case (proof of Prop. 5.4) there must exist a^i with $\bar{u}^i_i \geq c_i(a^i)$ and

$$G(a^i)(1-\delta) + \delta G(\bar{a}^e) - \bar{u}^i_i \geq \bar{u}^i_j \geq c_j(a^i)(1-\delta) + \delta \bar{u}^j_j \text{ for } i \neq j \in \{1,2\}.$$

Due to these conditions, there is a stationary contract $\sigma(\bar{a}^e, a^1, a^2)$ and punishment payoffs $\bar{u}^1_1$ and $\bar{u}^2_2$. In this stationary contract, all continuation equilibria either have total payoff $\bar{U}_{SGP}$, or the payoff is u^i with $u^i_i = \bar{u}^i_i$ and $u^i_j \geq \bar{u}^i_j$. Thus, $\sigma(\bar{a}^e, a^1, a^2)$ is strong perfect for all subgame perfect up-front payments. Therefore, $\mathcal{U}_{SP}^{pay}$ must be equal to the line from $(\bar{u}^1_1, G(\bar{a}^e) - \bar{u}^1_1)$ to $(G(\bar{a}^e) - \bar{u}^2_2, \bar{u}^2_2)$.

Proof of Proposition 5.8.
First we show that a stationary contract $\sigma(\bar{a}^e, a^1, a^2)$ with punishment payoffs u_1^1 and u_2^2 is strong perfect given that the condition in the proposition holds. Clearly, continuation equilibria that start with the transfer cannot be Pareto dominated, so we only have to show that this holds also for the punishment phases. Assume to the contrary that there is an equilibrium $\tilde{\sigma} \in \Sigma_{SGP}^{play}$ that strictly dominates the punishment for player i. The first action profile $\tilde{a}$ of $\tilde{\sigma}$ is admissible and since $G(a^i)(1-\delta) + \delta G(\bar{a}^e) < U(\tilde{\sigma}) \leq G(\tilde{a})(1-\delta) + \delta G(\bar{a}^e)$ it must hold that $G(\tilde{a}) > G(a^i)$, hence either inequality (i) or (ii) holds. We know that in the equilibrium $\tilde{\sigma}$ player j's payoff is bounded by the joint payoff $U(\tilde{\sigma})$ minus player i's minimum payoff $(1-\delta)c_i(\tilde{a}) + \delta\bar{u}_{SGP}^i$. Hence strict Pareto dominance of $\tilde{\sigma}$ implies that

$$(1-\delta)G(a^i)+\delta G(\bar{a}^e)-u_i^i < u_j(\tilde{\sigma}) \leq G(\tilde{a})(1-\delta)+\delta G(\bar{a}^e)-(1-\delta)c_i(\tilde{a})-\delta\bar{u}_{SGP}^i$$

and

$$u_i^i < u_i(\tilde{\sigma}) \leq G(\tilde{a})(1-\delta) + \delta G(\bar{a}^e) - (1-\delta)c_j(\tilde{a}) - \delta\bar{u}_{SGP}^j,$$

which is a contradiction to the fact that either (i) or (ii) has to hold.

Next we assume that $\sigma(\bar{a}^e, a^1, a^2)$ with punishment payoffs u_1^1, u_2^2 is strong perfect, and assume to the contrary that there exists an admissible $\tilde{a}$ with $G(\tilde{a}) > G(a^i)$ and

$$(1-\delta)G(\tilde{a}) - c_i(\tilde{a})(1-\delta) - \delta\bar{u}_{SGP}^i > (1-\delta)G(a^i) - u_i^i \tag{5.23}$$

and

$$(1-\delta)G(\tilde{a}) + \delta G(\bar{a}^e) - c_j(\tilde{a})(1-\delta) - \bar{u}_{SGP}^j > u_i^i. \tag{5.24}$$

for some player i. Because $\tilde{a}$ is admissible, strategies that follow play of the path

$$\tilde{Q} := (\tilde{a}, \tilde{p}, a^e, p^e, a^e, ...)$$

and use the optimal penal codes as punishments form a subgame perfect equilibrium for a nonempty range of $\tilde{p}$. Since $G(\tilde{a}) > G(a^i)$ the equilibrium has a higher joint payoff than the punishment phase for player i in our stationary equilibrium. Because of conditions (5.23) and (5.24), the transfer after play of $\tilde{a}$ can be used to give each player strictly more than in the punishment phase of $\sigma(\bar{a}^e, a^1, a^2)$, contradicting strong perfection.

Proof of Corollary 5.9.
Assume first that there exist strong perfect stationary contracts with punishment payoffs $\bar{u}_{SGP}^i$. There must exist an optimal stationary contract $\sigma(\bar{a}^e, \bar{a}^1, \bar{a}^2)$ with maximum fines and such that $c_i(\bar{a}^i) = \bar{u}_{SGP}^i$ and $G(\bar{a}^i) \geq G(a^i)$ for all other admissible action profiles a^i with $c_i(a^i) = \bar{u}_{SGP}^i$. Now assume that there is an admissible action $\tilde{a}$ and $i \in \{1, 2\}$ with $G(\bar{a}^i) - c_i(\bar{a}^i) <$

$G(\tilde{a})-c_i(\tilde{a})$. Since $c_i(\bar{a}^i) \leq c_i(\tilde{a})$ it must hold that $G(\tilde{a}) > G(a^i)$. Condition (i) of Prop. 5.8 then takes the form $G(\tilde{a}) - c_i(\tilde{a}) \leq G(\bar{a}^i) - c_i(\bar{a}^i)$ and therefore does not hold. Condition (ii), together with admissibility of $\tilde{a}$, implies that

$$c_i(\tilde{a})(1-\delta)+\delta c_i(\bar{a}^i) \leq G(\tilde{a})(1-\delta)+\delta G(\bar{a}^e) - c_j(\tilde{a})(1-\delta) - \delta c_j(\bar{a}^j) \leq c_i(\bar{a}^i).$$

Hence, $c_i(\tilde{a}) = c_i(\bar{a}^i)$, contradicting our choice of $\bar{a}^i$.

Next, assume that there are admissible $\bar{a}^1$ and $\bar{a}^2$ such that for all admissible $\tilde{a}$ it holds that

$$G(\bar{a}^i) - c_i(\bar{a}^i) \geq G(\tilde{a}) - c_i(\tilde{a}) \text{ for } i = 1, 2.$$

Then there exists an optimal stationary contract $\sigma(\bar{a}^e, \bar{a}^1, \bar{a}^2)$ with $c_i(\bar{a}^i) = \bar{u}^i_{SGP}$, and condition (i) of Prop. 5.8 is true for all admissible $\tilde{a}$.

Proof of Corollary 5.10.
Assume to the contrary that there is a strong perfect stationary contract $\sigma(\bar{a}^e, a^1, a^2)$ with punishment payoffs u^1_1 and u^2_2, while it holds that

$$(1-\delta)\left(G(\bar{a}^e) - c_i(\bar{a}^e)\right) - \delta c_i(\bar{a}^i) > (1-\delta)G(a^i) - c_i(a^i) \text{ for both } i \in \{1, 2\}.$$

Take $\tilde{a} = \bar{a}^e$ in Prop. 5.8. Condition (i) does not hold for any $i \in \{1, 2\}$. Therefore, condition (ii) must hold for both $i \in \{1, 2\}$, and in sum these conditions imply that $u^1_1 + u^2_2 \geq G(\bar{a}^e)$. This can only be fulfilled if $u^1_1 + u^2_2 = G(\bar{a}^e)$, and in this case the condition also implies that $G(\bar{a}^e) = c_1(\bar{a}^1)+c_2(\bar{a}^2)$. This means that $\mathcal{P}(\mathcal{U}^{pay}_{SGP})$ and $\mathcal{P}(\mathcal{U}^{play}_{SGP})$ consist of at most one point with joint payoff $\bar{U}_{SGP}$, hence $\sigma(\bar{a}^e, a^1, a^2)$ cannot be strong perfect.

Proof of Lemma 5.14.
Let $\sigma(a^e, a^1, a^2)$ with equilibrium payoff u^e be a regular stationary contract that is WRP. If $(1-\delta)G(a^i) + \delta G(a^e) - c_i(a^i) < u^e_j$, then it must hold that $c_i(a^i) \geq u^e_i$, i.e., $c_i(a^i) \geq c_i(a^e)$. This implies $a^i = a^e$ and therefore

$$G(a^i)(1-\delta) + \delta G(a^e) - c_i(a^i) = G(a^e) - c_i(a^e) \geq G(a^e) - u^e_i = u^e_j.$$

Next, assume that for the regular stationary contract $\sigma(a^e, a^1, a^2)$ inequality (5.10) holds. Since $G(a^e) \geq G(a^i)$ this implies that the payoff when player i is punished and the equilibrium payoff u^e cannot be Pareto ranked. Moreover, $(1-\delta)G(a^i) + \delta G(a^e) - c_i(a^i) \geq u^e_j \geq c_i(a^i)$, and therefore the two punishments cannot be Pareto ranked, either.

To prove the second statement, note that the conditions (5.11) and (5.12) follow from the condition 5.10 and subgame perfection. To prove the other direction, one has to adjust the equilibrium payment in the appropriate way.

Conditions (5.11) and (5.12) and the conditions for subgame perfection imply that it is possible to choose $\tilde{p}_1^e$ such that

$$G(a^e) - c_2(a^e)(1-\delta) - \delta c_2(a^2) \geq g_1(a^e) - \tilde{p}_1^e \geq c_1(a^e)(1-\delta) + \delta c_1(a^1)$$

and

$$G(a^2)(1-\delta) + \delta G(a^e) - c_2(a^2) \geq g_1(a^e) - \tilde{p}_1^e \geq (1-\delta)(G(a^e) - G(a^1)) - c_1(a^1)$$

The resulting stationary contract, which is equal to $\sigma(a^e, a^1, a^2)$ except for the new equilibrium payment $\tilde{p}^e$, is still a regular stationary contract, and it is WRP.

Proof of Proposition 5.15.
Let σ be any WRP equilibrium and let $\bar{U} = \sup_{u \in \mathcal{U}^{play}(\sigma)} u_1 + u_2$, and $\bar{u}_i^i = \inf_{u \in \mathcal{U}^{play}(\sigma)} u_i$. We take $(\bar{u}_1^e, \bar{u}_2^e)$ to be a payoff tuple in the closure of $\mathcal{U}^{play}(\sigma)$ such that $\bar{u}_1^e + \bar{u}_2^e = \bar{U}$. Similarly, $(\bar{u}_1^i, \bar{u}_2^i)$ shall be a tuple in the closure of $\mathcal{U}^{play}(\sigma)$ such that among all such tuples with the same player i payoff $\bar{u}_i^i$, player j's payoff is maximized. We then have that $\bar{u}_i^i \leq u_i$ and $\bar{u}_j^i \geq u_j$ for all $u \in \mathcal{U}^{play}(\sigma)$. Let $u(\sigma|h_n^e)$ be a sequence in $\mathcal{U}^{play}(\sigma)$ with limit $(\bar{u}_1^e, \bar{u}_2^e)$ and for $i = 1, 2$ let $u(\sigma|h_n^i)$ be a sequence with limit $(\bar{u}_1^i, \bar{u}_2^i)$. Let furthermore a_n^l be the w.l.o.g. convergent sequences of the first action profiles of the continuation equilibria $\sigma|h_n^l$, $l = e, 1, 2$. Completely analogous to the SGP case (see the proof of Prop. 5.4) we have for the limits of these sequences, denoted by a^e, a^1, a^2, that $G(a^e) \geq \bar{U}$, $c_i(a^i) \leq \bar{u}_i^i$,

$$\bar{U} \geq (c_1(a^e) + c_2(a^e))(1-\delta) + \delta(\bar{u}_1^1 + \bar{u}_2^2),$$

and

$$G(a^i)(1-\delta) + \delta\bar{U} - \bar{u}_i^i \geq \bar{u}_j^i \geq \max(c_j(a^j), c_j(a^i))(1-\delta) + \delta\bar{u}_j^j,$$

as well as

$$G(a^i)(1-\delta) + \delta\bar{U} - c_i(a^i) \geq \bar{u}_j^e \geq c_j(a^e)(1-\delta) + \delta c_j(a^j)$$

and

$$(G(a^1) + G(a^2))(1-\delta) + 2\delta\bar{U} - c_1(a^1) - c_2(a^2) \geq \bar{U}.$$

Since we assumed that $\delta \geq \frac{1}{2}$, these conditions are relaxed if we replace $\bar{U}$ by $G(a^e)$. Next, define $\tilde{a}^e \in \{a^e, a^1, a^2\}$ such that $G(\tilde{a}^e) = \max\{G(a^e), G(a^1), G(a^2)\}$, and $\tilde{a}^i = a^i$ if $c_i(a^i) < c_i(\tilde{a}^e)$ and $\tilde{a}^i = \tilde{a}^e$ else. It is straightforward to show that all conditions still hold:

$$G(\tilde{a}^e) \geq (1-\delta)(c_1(\tilde{a}^e) + c_2(\tilde{a}^e)) + \delta(c_1(\tilde{a}^1) + c_2(\tilde{a}^2)),$$

$$G(\tilde{a}^i)(1-\delta) + \delta G(\tilde{a}^e) - c_i(\tilde{a}^i) \geq \max(c_j(\tilde{a}^e), c_j(\tilde{a}^i))(1-\delta) + \delta c_j(\tilde{a}^j),$$

and

$$(G(\tilde{a}^1) + G(\tilde{a}^2))(1-\delta) + 2\delta G(\tilde{a}^e) - (c_1(\tilde{a}^1) + c_2(\tilde{a}^2)) \geq G(\tilde{a}^e).$$

Because of Lemma 5.14 there is a WRP regular stationary contract $\sigma(\tilde{a}^e, \tilde{a}^1, \tilde{a}^2)$ with properties as stated in the proposition.

Proof of Proposition 5.16.
Since no two payoff tuples in $\mathcal{U}^{play}_{SRP}$ can be strictly Pareto ranked, one can show as in the WRP case that there exists a WRP regular stationary contract $\sigma(a^e, a^1, a^2)$ with $G(a^e) \geq u_1 + u_2$, $c_i(a^i) \leq u_i$,and $G(a^i)(1-\delta) + \delta G(a^e) - c_i(a^i) \geq u_j$ for all $u \in \mathcal{U}^{play}_{SRP}$. Since $\sigma(a^e, a^1, a^2)$ cannot Pareto dominate the SRP equilibria it follows that $G(a^e) = \max_{\mathcal{U}^{play}_{SRP}} u_1 + u_2$, and because the worst SRP payoffs must be able to sustain a^e, $\sigma(a^e, a^1, a^2)$ with punishment payoffs $\inf_{\mathcal{U}^{play}_{SRP}} u_i$ instead of $c_i(a^i)$ is SRP.

Proof of Proposition 5.17.
Assume that $\sigma(a^e, a^1, a^2)$ is not SRP. Since $\sigma(a^e, a^1, a^2)$ is an optimal WRP stationary contract, it can only be dominated in the punishment phase, that is, there must be $i \in \{1, 2\}$ and a WRP equilibrium $\tilde{\sigma}$ such that $u_i(\tilde{\sigma}|h) > c_i(a^i)$ and for $j = 3 - i$

$$u_j(\tilde{\sigma}|h) > G(a^i)(1-\delta) + \delta G(a^e) - c_i(a^i)$$

for some $h \in H^{play}$. Because of Prop. 5.15 there exists a regular WRP stationary contract $\sigma(\tilde{a}^e, \tilde{a}^1, \tilde{a}^2)$ with

$$G(\tilde{a}^i)(1-\delta) + \delta G(\tilde{a}^e) - c_i(\tilde{a}^i) \geq u_j(\tilde{\sigma}|h).$$

Proof of Proposition 5.19.
The proof is analogous to the proof of Proposition 5.15. Let $\sigma \in \Sigma^{pay}_{WRP}$ and as before, let $u(\sigma|h^e_n)$ be a sequence in $\mathcal{U}^{play}(\sigma)$ with limit $\bar{u}^e$ such that $\bar{u}^e_1 + \bar{u}^e_2 = \bar{U}^e = \sup_{u \in \mathcal{U}^{play}(\sigma)} u_1 + u_2$, and let $u(\sigma|h^i_n)$ be a sequence with limit $\bar{u}^i$, where $\bar{u}^i_i \leq u_i$ and $\bar{u}^i_j \geq u_j$ for all $u \in \mathcal{U}^{play}(\sigma)$. Let furthermore a^e, a^1, a^2 be the limits of the sequences consisting of the first action pairs of the continuation equilibria $\sigma|h^l_n$, $l = e, 1, 2$. We already know that $G(a^e) \geq \bar{U}^e$,

$$G(a^i)(1-\delta) + \delta\bar{U}^e - c_i(a^i) \geq \max(c_j(a^i), c_j(a^j))(1-\delta) + \delta c_j(a^j),$$
$$G(a^i)(1-\delta) + \delta\bar{U}^e - c_i(a^i) \geq \bar{u}^e_j \geq c_j(a^e)(1-\delta) + \delta c_j(a^j),$$

and in particular

$$(G(a^1) + G(a^2))(1-\delta) + (2\delta - 1)\bar{U}^e - c_1(a^1) - c_2(a^2) \geq 0. \tag{5.25}$$

If this last condition still holds if we replace $\bar{U}^e$ by $G(a^e)$, then we are in the case of Prop. 5.15. For the case that $c_i(a^i) \geq c_j(a^e)$ for some i, it holds that $\sigma(a^e, a^e, a^j)$ is a WRP stationary contract.

Let us now assume that $c_i(a^i) < c_i(a^e)$ and 5.25 binds at some joint payoff $\bar{U}^e \leq \bar{U} < G(a^e)$. This payoff is the one defined in (5.14), and the conditions for subgame perfection (5.15) are satisfied as well. Therefore, we know that there is a stationary contract $\sigma(a^e, a^1, a^2)$ with money burning and equilibrium payoff $\big(G(a^1)(1-\delta) + \delta\bar{U}^e - c_1(a^1), G(a^2)(1-\delta) + \delta\bar{U}^e - c_2(a^2)\big)$. The WRP conditions for the play stage hold by definition, it remains only to show that they also hold for the pay stage.

We show first that no tuple in the closure of $\mathcal{U}^{pay}(\sigma)$ strictly Pareto dominates another such tuple: take any two sequences $u(\sigma|h_n)$ and $u(\sigma|h'_n)$ in $\mathcal{U}^{pay}(\sigma)$. Then for any $n \in \mathbb{N}$ it holds that either $u_1(\sigma|h_n) \geq u_1(\sigma|h'_n)$ and $u_2(\sigma|h_n) \leq u_2(\sigma|h'_n)$, or $u_1(\sigma|h_n) \leq u_1(\sigma|h'_n)$ and $u_2(\sigma|h_n) \geq u_1(\sigma|h'_n)$. Since one of the conditions must hold for a subsequence, this condition must then also hold for the limit.

Consequently, among the payoff vectors $u(\sigma)$, $(\bar{u}^e - g(a^e)(1-\delta))/\delta$, and $(\bar{u}^i - g(a^i)(1-\delta))/\delta$, $i = 1, 2$ no two are strictly Pareto ranked. This implies that $\bar{U} \geq g_i(a^e) - g_j(a^i)$ or $g_j(a^e) \leq g_j(a^i)$. Moreover, there must be an $i \in \{1, 2\}$ with $\delta u_i(\sigma) \leq \bar{u}^e_i - g_i(a^e)(1-\delta)$. Define the up-front transfer in $\sigma(a^e, a^1, a^2)$ such that $u_i(\sigma(a^e, a^1, a^2)) = u_i(\sigma)$, the resulting $\sigma(a^e, a^1, a^2)$ then is a money-burning WRP stationary contract that weakly Pareto dominates σ.

References

Abreu, Dilip, "Extremal Equilibria of Oligopolistic Supergames," *Journal of Economic Theory*, 1986.

——, "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 1988, *56*, 383–396.

Aghion, Philippe, Mathias Dewatripont, and Patrick Rey, "Renegotiation Design with Unverifiable Information," *Econometrica*, 1994, *62* (2), 257–82.

Akerlof, George A. and Lawrence F. Katz, "Workers' Trust Funds and the Logic of Wage Profiles," *Quarterly Journal of Economics*, 1989, *104*, 525–536.

Andersson, Ola and Erik Wengstrm, "A Note on Renegotiation in Repeated Bertrand Duopolies," *Economics Letters*, 2007, *95* (3), 398–401.

Baker, George, Robert Gibbons, and Kevin J. Murphy, "Relational Contracts and the Theory of the Firm," *The Quarterly Journal of Economics*, 2002, *2*, 39–84.

Baliga, Sandeep and Robert Evans, "Renegotiation in Repeated Games with Side Payments," *Games and Economic Behavior*, 2000, *33*, 159–176.

—— **and Tomas Sjöström**, "Decentralization and Collusion," *Journal of Economic Theory*, 1998, *83*, 196–232.

Blonski, Matthias and Giancarlo Spagnolo, "Relational Contracts and Property Rights," *working paper*, 2003.

Bolton, Patrick and Mathias Dewatripont, *Contract Theory*, Cambridge,MA: MIT Press, 2005.

Brennan, Jim and Joel Watson, "The Renegotiation-Proofness Principle and Costly Renegotiation," *UCSD Working Paper*, 2002.

Che, Yeon-Koo and Seung-Weon Yoo, "Optimal Incentives for Teams," *American Economic Review*, 2001, *91*, 525–541.

Chiappori, Pierre, Ines Macho, Patrick Rey, and Bernard Salanié, "Repeated Moral Hazard: The Role of Memory Commitment and the Access to Credit Markets," *European Economic Review*, 1994, *38*, 1527–1553.

Chonne, Philippe and Laurent Linnemer, "Optimal Litigation Strategies with Signaling and Screening," *working paper*, 2008.

Chung, Tai-Yeong, “Incomplete Contracts, Specific Investments, and Risk Sharing,” *Review of Economic Studies*, 1991, *58* (5), 1031–42.

Coase, Ronald H., “The Nature of the Firm,” *Economica*, 1937, *4* (16), 386405.

Cooter, Robert, “Unity in Tort, Contract and Property: the Model of Precaution,” *California Law Review*, 1985, *73* (1), 1–51.

Crémer, Jacques, “Arm’s Length Relationships,” *Quarterly Journal of Economics*, 1995, *110*, 275–295.

Daughety, Andrew F. and Jennifer F. Reinganum, “Hidden Talents: Partnerships with Pareto-Improving Private Information,” 2009. working paper.

de Meza, David and Ben Lockwood, “Does Asset Ownership Always Motivate Managers? Outside Options and the Property Rights Theory of the Firm,” *The Quarterly Journal of Economics*, 1998, pp. 361–386.

Dewatripont, Mathias and Eric Maskin, “Contract Renegotiation in Models of Asymmetric Information,” *European Economic Review*, 1990, *34*, 311–321.

___ **and** ___ , “Contractual Contingencies and Renegotiation,” *The RAND Journal of Economics*, 1995, *26*, 704–719.

___ **and** ___ , “Credit and Efficiency in Centralized and Decentralized Economies,” *Review of Economic Studies*, 1995, *62*, 541–555.

___ **, Patrick Legros, and Steven A. Matthews**, “Moral Hazard and Capital Structure Dynamics,” *Journal of the European Economic Association*, 2003, *1*, 890–930.

Doornik, Katherine, “Relational Contracting in Partnerships,” *Journal of Economics & Management Strategy*, 2006, *15* (2).

Edlin, Aaron S., “Cadillac Contracts and Up-front Payments: Efficient Investments under Expectation Damages,” *Journal of Law, Economics, and Organization*, 1996, *12* (1), 98–118.

___ **and Stefan Reichelstein**, “Holdups, Standard Breach Remedies, and Optimal Investments,” *American Economic Review*, 1996, *86* (3), 478–501.

Evans, Robert, “Simple Efficient Contracts in Complex Environments,” *Econometrica*, 2008, *76* (3), 459–491.

Farrell, Joseph and Eric Maskin, “Renegotiation in Repeated Games,” *Games and Economic Behavior*, 1989, *1*, 327–360.

___ **and Georg Weizsacker**, “Renegotiation in the Repeated Amnesty Dilemma, with Economic Applications,” in Kalyan Chatterjee and William F. Samuelson, eds., *Game Theory and Business Applications*, Vol. 35 of *International Series in Operations Research & Management Science*, Springer US, 2001, pp. 213–246.

___ **and Robert Gibbons**, “Cheap Talk about Specific Invesments,” *Journal of Law, Economcis, and Organization*, 1995, *11* (2), 313–334.

Fong, Yuk fai and Jay Surti, “On the Optimal Degree of Cooperation in the Repeated Prisoner’s Dilemma witht Side Payments,” *Games and Economic Behavior, forthcoming*, 2009.

Fudenberg, Drew and Jean Tirole, “Moral Hazard and Renegotiation in Agency Contracts,” *Econometrica*, 1990, *58*, 1279–1319.

___ **, Bengt Holmström, and Paul Milgrom**, “Short-Term Contracts and Long-Term Agency Relationships,” *Journal of Economic Theory*, 1990, *51*, 1–31.

Gilovich, T., R. Vallone, and A. Tversky, "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 1985, *17*, 295–314.

Gonzales, Patrick, "Investment and Screening under asymmetric endogenous information," *RAND Journal of Economics*, 2004, *35* (3), 502–519.

Grossman, Sanford J. and Oliver Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Obligation," *Journal of Political Economy*, 1986, *94*, 691–719.

Grout, Paul A., "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," *Econometrica*, 1984, *52* (2), 449–460.

Gul, Faruk, "Unobservable Investment and the Hold-up Problem," *Econometrica*, 2001, *69* (2), 343–376.

Guriev, Sergei and Dmitriy Kvasov, "Contracting on Time," *American Economic Review*, 2005, *95* (5), 1369–85.

Harrington, Joseph E., "How do Cartels Operate?," *Foundations and Trends in Microeconomics*, 2006, *2* (1), 1–105.

___ **and Andrzej Skrzypacz**, "Collusion under Monitoring of Sales.," *The RAND Journal of Economics*, 2007.

Hart, Oliver and John Moore, "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 1990, *98* (6), 1119–1158.

Hart, Oliver D., *Firms, Contracts, and Financial Structure*, Clarendon Press Oxford, 1995.

___ **and John H. Moore**, "Incomplete Contracts and Renegotiation," *Econometrica*, 1988, *56* (4), 755–785.

Hermalin, Benjamin E., "Toward an Economic Theory of Leadership: Leading by Example," *American Economic Review*, 1998, *88* (5), 1188–1206.

___ **and M.L Katz**, "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency," *Econometrica*, 1991, *59*, 1735–1753.

Innes, Robert D., "Limited Liability and Incentive Contracting with Ex-ante Action Choices," *Journal of Economic Theory*, 1990, *52*, 45–67.

Jullien, Bruno, "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 2000, *93* (1), 1–47.

Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *The Journal of Law and Economics*, 1978, *21*, 297–326.

Kornai, János., Eric S. Maskin, and Gérard Roland, "Understanding the Soft Budget Constraint," *Journal of Economic Literature*, 2003, *41*, 1095–1136.

Kranz, Sebastian and Susanne Ohlendorf, "Renegotiation-proof relational contracts with side payments," May 2009. SFB TR 15 discussion paper No. 259.

Laffont, Jean-Jaques and David Martimort, *The Theory of Incentives: The Principal-Agent Model*, Princeton, N.J.: Princeton University Press, 2002.

Lazear, Edward P., "Agency, Earnings Profiles, Productivity, and Hours Restrictions," *American Economic Review*, 1981, *71*, 606–620.

Leitzel, Jim, "Damage Measures and Incomplete Contracts," *RAND Journal of Economics*, 1989, *20* (1), 92–101.

Levin, Jonathan, "Relational Incentive Contracts," *American Economic Review*, 2003, *93*, 835–857.

Lewis, Tracy R. and David E.M. Sappington, "Choosing Workers' Qualifications: No Experience Necessary?," *International Economic Review*, 1993, *34* (3), 479–502.

___ **and** ___ , "Contracting With Wealth-Constrained Agents," *International Economic Review*, 2000, *41*, 743–767.

Ma, Ching-to Albert, "Adverse Selection in Dynamic Moral Hazard," *The Quarterly Journal of Economics*, 1991, *106* (1), 255–275.

___ , "Renegotiation and Optimality in Agency Contracts," *Review of Economic Studies*, 1994, *61*, 109–129.

Macaulay, Stewart, "Non-contractual Relations in Business: A Preliminary Study," *American Sociological Review*, 1963, *28* (1), 55–67.

MacLeod, W. Bentley, "Reputations, Relationships, and Contract Enforcement," *Journal of Economic Literature*, 2007, *45*, 595–628.

MacLeod, William Bentley and James M. Malcomson, "Investments, Holdup, and the Form of Market Contracts," *American Economic Review*, 1993, *83* (4), 811–37.

Macneil, Ian R., "Contracts: Adjustment of Long-term Economic Relations under Classical, Neoclassical, and Relational Contract Law," *Northwestern University Law Review*, 1978, *72* (6), 854–905.

Malcomson, James M., "Contracts, hold-up, and labor markets," *Journal of Economic Literature*, 1997, *35* (4), 1916–1957.

___ **and Frans Spinnewyn**, "The Multiperiod Principal-Agent Problem," *Review of Economic Studies*, 1988, *55*, 391–407.

Matthews, Steven A., "Renegotiation of Sales Contracts," *Econometrica*, 1995, *63*, 567–591.

___ , "Renegotiating Moral Hazard Contracts under Limited Liability and Monotonicity," *Journal of Economic Theory*, 2001, *97*, 1–29.

McCutcheon, Barbara, "Do Meetings in Smoke-Filled Rooms Facilitate Collusion?," *Journal of Political Economy*, 1997, *105* (2), 330–350.

McFall, Todd A., C.R. Knoeber, and W.N. Thurman, "Contest, Grand Prizes, and the Hot Hand," *Discussion Paper*, 2006.

Miklos-Thal, Jeanine, "Optimal Collusion under Cost Asymmetry," *working paper*, 2008.

Moore, John, "Optimal Labour Contracts when Workers have a Variety of Privately Observed Reservation Wages," *The Review of Economic Studies*, 1985, *52* (1), 37–67.

Nöldeke, Georg and Klaus M. Schmidt, "Option Contracts and Renegotiation: A Solution to the Holdup Problem," *RAND Journal of Economics*, 1995, *26* (2), 163–79.

Ohlendorf, Susanne, "Expectation Damages, Divisible Contracts, and Bilateral Investment," *American Economic Review*, forthcoming.

___ **and Patrick W. Schmitz**, "Repeated Moral Hazard, Limited Liability, and Renegotiation," February 2008. CEPR discussion paper No. 6725.

Pitchford, Rohan, "Moral Hazard and Limited Liability: The Real Effects of Contract Bargaining," *Economics Letters*, 1998, *61*, 251–259.

Rey, Patrick and Bernard Salanié, "Long-Term, Short-Term and Renegotiation: On the Value of Commitment in Contracting," *Econometrica*, 1990, *58*, 597–619.

Riley, John G., "Silver Signals: Twenty-Five Years of Screening and Signaling," *Journal of Economic Literature*, 2001, *39*, 432–478.

Rockafellar, R. Tyrrell, *Convex Analysis*, Princeton University Press, 1970.

Rogerson, William P., "Efficient Reliance and Damage Measure for Breach of Contract," *RAND Journal of Economics*, 1984, *15* (1), 39–53.

___ , "Repeated Moral Hazard," *Econometrica*, 1985, *53*, 69–76.

Schmitz, Patrick W., "Allocating Control in Agency Problems with Limited Liability and Sequential Hidden Actions," *RAND Journal of Economics*, 2005, *36*, 318–336.

Schweizer, Urs, "The Pure Theory of Multilateral Obligations," *Journal of Institutional and Theoretical Economics*, 2005, *161* (2), 239–254.

Segerstrom, Paul S., "Demons and Repentence," *Journal of Economic Theory*, 1988, *45* (1), 32–52.

Shapiro, Carl and Joseph Stiglitz, "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 1984, *74*, 433–444.

Shavell, Steven, "Damage Measures for Breach of Contract," *Bell Journal of Economics*, 1980, *11* (2), 466–490.

Skrzypacz, Andrzej, "Bargaining under Asymmetric Information and the Hold-up Problem," April 2004. working paper.

Sloof, Randolph, "Price-setting power vs. private information: An experimental evaluation of their impact on holdup," *European Economic Review*, 2008, *52* (3), 469–486.

Spence, A. Michael, "Job Market Signaling," *Quarterly Journal of Economics*, 1973, *87* (3), 355–379.

Strausz, Roland, "Buried in Paperwork: Excessive Reporting in Organizations," *Journal of Economic Behavior and Organization*, 2006, *60*, 460–470.

Tirole, Jean, "Incomplete Contracts: Where do we Stand?," *Econometrica*, 1999, *67* (4), 741–781.

___ , "Corporate Governance," *Econometrica*, 2001, *69*, 1–35.

___ , *The Theory of Corporate Finance*, Princeton, N.J.: Princeton University Press, 2005.

van Damme, Eric, "Renegotiation-Proof Equilibria in Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 1989, *47*, 206–217.

Von Siemens, Ferdinand, "Bargaining Under Incomplete Information, Fairness, and the Hold-Up Problem," *SSRN eLibrary*, March 2007.

Weiss, Andrew, "A Sorting cum-learning Model of Education," *Journal of Political Economy*, 1983, *91* (3), 420–442.

Williamson, Oliver E., "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, 1971, *61*, 112–123.

___ , "Transaction Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 1979, *22*, 233–261.

___ , *The Economic Institutions of Capitalism*, New York: The Free Press, 1985.