Essays in Microeconomics

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Diplom-Math. Martin Pollrich

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Dr. Ulrich Kamecke

Gutachter:
1. Prof. Dr. Roland Strausz
2. Prof. Paul Heidhues PhD.
3. Prof. Dr. Elmar Wolfstetter

Eingereicht am: 15.10.2014
Tag des Kolloquiums: 30.04.2015

Contents

List of Figures

Introduction

It would be presumptuous to claim that the three chapters, which make up this thesis, are closely related. Broadly categorized the first chapter originates from the area of industrial organization, the second chapter contributes to the theory of contracts, and the third chapter to the economics of regulation. However, a closer look unveils a common theme: the role of commitment, specifically the absence of commitment.

The Oxford Dictionary defines *commitment* as "an engagement or obligation that restricts freedom of action".[1] In economics we are trained to think of any interaction within a group of individuals as a game. Using this understanding, commitment means that an individual promises to refrain from certain actions during play. The benefits and drawbacks of commitment can be best understood with the following example.

Consider parents trying to teach their child the proper use of cutlery through trial and error.[2] Committing to sufficiently long periods of experimentation, without interference, is acknowledged as an appropriate strategy. The child's errors, however, can undermine the parents' credibility. When the child learns that parents will help with eating, once the mess created on and around the table is sufficient, it will learn how to induce the parents' help faster than how to use the spoon.

The example highlights the importance of credibility for the usefulness of making commitments. The three chapters of this thesis study similar cases, where the assumption of commitment to particular actions is questioned. They aim to understand the impact of relaxing such an assumption, how economic outcomes are affected and which conclusions concerning the involved institutions can be drawn.

Chapter 1 is based on joint work with Lilo Wagner (Pollrich and Wagner (2014)). It studies the interaction of information disclosure and reputational concerns in certi-

[1] The definition is taken from http://www.oxforddictionaries.com/definition/english/commitment. It refers to the second definition of the word commitment.

[2] To translate this example into a purely economic environment, think of the parents as a principal and the child as the agent. Learning the use of cutlery can then be replaced by an arbitrary task, such as instructing the agent in the use of a new machine during production.

fication markets. In this chapter we relax the assumption that a certifier can commit to a disclosure rule, and instead allow for the possibility that sellers bribe the certifier to get favorable certificates. We argue that revealing less precise, resp. coarse, information helps the certifier maintain credibility, and turn down bribes.

Underlying this effect is a classic trade-off the certifier faces in his long-run relationship with consumers. He trades-off short-term gains from bribes, versus the long-term losses from losing trust, which boil down to losing any future profit. Opaque disclosure rules constrain feasible bribes, thus, lowering the potential short-run gain. On the other hand, long-run losses are - if at all - only slightly reduced. As a consequence, only less precise disclosure rules are implementable for intermediate discount factors. These insights suggest that contrary to the common view, coarse disclosure is socially desirable. A ban may provoke market failure, especially in industries where certifier reputational rents are low.

In the second chapter I study the optimal communication mechanism in a setting with adverse selection when the principal cannot commit to an audit strategy. Usually, the revelation principle applied to principal-agent problems with full commitment implies that the principal can offer a menu of contracts containing one item per type of agent. The principal's inability to credibly commit to an audit strategy leads to the undermining of this simple structure. Rather, the principal resorts to an impartial mediator.

The optimal mediated contract has the following properties: (1) the agent reports truthfully to the mediator, (2) the mediator performs a report-dependent randomization, (3) the randomization is over a report-dependent transfer-quantity schedule and a fixed transfer-quantity schedule, (4) only the fixed transfer-quantity schedule is accompanied with a recommendation to audit, and (5) the principal obediently follows the mediator's recommendation. Properties (1) and (5) follow from applying the revelation principle. (2) is used to guarantee the principal's obedience after a recommendation to audit. In particular, it creates the right posterior that makes the principal indifferent between auditing and not when the recommended action is audit. The exact details of the three utilized transfer-quantity schedules is determined by trading off rents, efficiency and audit costs.

It is crucial to employ a mediator for the randomization, otherwise the principal learns the agent's type from his report and updates her belief accordingly. Intuitively, the mediator breaks the information flow and transmits only as much information to the principal as is required to guarantee obedience. The structure of the optimal mechanism improves our understanding of the institutional design of audit agencies. A separation of auditing and contracting is inherent in the optimal mechanism and should therefore also be reflected in the institutions. Separate audit agencies are just one way of implementing this idea.

Chapter 3 is based on joint work with Robert Schmidt (Pollrich and Schmidt (2014)). It studies the design of optimal contracts to avert firm relocation. One way that a firm's relocation can be triggered, is a unilateral policy intervention such as the introduction of an emission price. We analyze a dynamic game where a regulator offers contracts to avert the relocation of a firm in each of two periods. The firm can undertake a location-specific investment (e.g., in abatement capital) at the beginning of the first period. Contracts can be written on some contractible productive activity (e.g., emissions), but the firm's investment is not contractible.

We analyze optimal contracts in this setting under two differing assumptions on the principal's ability to make commitments for future periods. When the regulator can commit to long-term contracts, simple subsidy payments are sufficient to avert relocation. However, an implementation problem arises when only short-term contracts are feasible. Because the second-period contract can only compensate the firm for not relocating in that period, any compensation for the firm's investment needs to be paid in period one. This, however, opens up the possibility for a 'take-the-money-and-run'-strategy for the firm: sack the large first-period transfer but secretly underinvest and relocate in period two. To prevent this the regulator resorts to high-powered incentives in the first period. The firm's investment is then so high that a lock-in effect prevents relocation in both periods. A second-period contract is, hence, no longer required.

Compared to the optimal contract under long-term contracting the firm now invests more and, consequently, receives a larger total transfer. Paradoxically, the distortion in the first-period contract can be so severe that higher transfers are needed to avert relocation compared to a (hypothetical) situation without the investment opportunity.

Chapter 1

Informational Opacity and Honest Certification

This chapter is based on Pollrich and Wagner (2014).

1.1 Introduction

In markets that exhibit informational asymmetries, product quality is typically reduced. This in turn may provoke a breakdown of trade. The lack of credible communication between informed and uninformed parties may result in the emergence of certification intermediaries. Certifiers inspect products whose characteristics are private information to agents, and publicly reveal this information. Examples of certifiers abound: rating agencies, eco-labels, wine certificates or technical inspections. Often however, certification results are revealed on a coarse scale, although the information at hand allows for a more precise disclosure.

A rich literature starting with Lizzeri (1999) has identified profit concerns as the cause for information disclosure by intermediaries being imprecise.[1] The rough intuition is that for the certifier's profit it is more important that many certificates are sold and not what the informational content of certificates is.

This chapter provides a novel explanation for such opacity: partially revealing rules can serve as a safeguard against fraud. Certifiers may be tempted to accept bribes for releasing favorable certificates, a behavior called capture. If consumers are aware of this threat of capture, the certifier must find ways to credibly commit to honesty.

[1] See the literature review at the end of this section.

The certifier faces a classic reputation dilemma in deciding whether the short-run gain from capture in form of the bribe is larger than the future profit losses from losing trust. We show that the certifier's choice of the disclosure rule has a crucial impact on this trade-off.

The effectiveness of employing opaque disclosure rules to signal credibility is independent of the motive of mere profit-maximization. In particular, opacity is beneficial even though certifier profits are maximized via revealing precise information. Consequently, coarse information revelation is also a feature of certification markets where profit concerns play only a minor role.[2]

Opacity in certification markets is yet another instance, where the commonly held view that reducing informational asymmetries is socially desirable per se, fails to be accurate.[3] Revealing too precise information is not credible to consumers and only coarse disclosure enhances information revelation in the first place. Hence, opacity can be welfare enhancing for the simple reason that it may prevent market failure.

Our results are important in the light of recent policy debates regarding the regulation of the market for credit ratings. The 'Dodd-Frank Wall Street Reform and Consumer Protection Act' includes without limitation regulations regarding the disclosure practice of rating agencies.[4] Though this kind of regulation would be innocent, if opacity was only caused by certifiers' profit concerns, it has a potential downside when the mere existence of the market depends on opacity. As we argue in this chapter, transparency is vulnerable against collusion between certifiers and sellers who demand certification. The economic purpose of opacity is to make the certification market work in the first place and not only to maximize profits.

We show our results in a model that allows us to delineate reputational and profit concerns - thus opacity caused by reputational concerns is present even when this does not maximize profits. We consider an infinitely repeated certification game with moral hazard where, in each period, short-lived producers first have to make an investment choice, which in turn determines the probability distribution of their product's quality. Thus, the payoffs assigned to each quality outcome determine the incentives to invest.

[2]Many non-profit organizations, such as the Marine Stewardship Council (MSC) or the Forest Stewardship Council (FSC), certify on a coarse scale, despite collecting fairly rich data on their clients.

[3]This point is, among others, made by Mason (2011), who argues that the introduction of noisy eco-labels may reduce welfare. Similarly, Langinier and Babcock (2008) study welfare effects when firms can certify as a group. Kreps and Wilson (1982) show that noise enhances welfare in finitely repeated games.

[4]E.g. Title IV, Sec. 404 and Sec. 405. For a comprehensive review of this Act and its impact on rating agencies see also Kartasheva and Yilmaz (2013)

The long-lived certifier has two instruments at his disposal: a flat certification fee and the disclosure rule. Consumers experience the true quality of a product only after consumption. If it does not match the awarded certificate, capture is detected. This makes the certifier face a classical reputation dilemma: she trades off short-run gains from capture against future profits.

We characterize feasible disclosure rules in this setting. Our major finding is that for sufficiently low discount factors, honest certification requires partial disclosure of quality information, which in our model implies noisy disclosure. In the short run, a certifier may gain from making a capture offer that is acceptable for at least some producers. The maximum producer willingness to pay for bribes is directly affected by the publicly announced disclosure rule. It is greatest for full disclosure and can be substantially reduced by revealing less precise information. But if consumers detect a bribe and therefore lose trust, a certifier gives up his future profits. Static certifier profits are maximal for full disclosure and any deviation will typically reduce the long-run loss from losing credibility. As will be shown, the first effect exceeds the latter.

We moreover obtain the counterintuitive result that a threat of capture increases social welfare.[5] Whenever information is fully revealed, sharing profits necessarily reduces producer investments as compared to the first-best level, obtained under complete information. We show that whenever capture offers are made before a certifier observes the true quality level, these are such that they are accepted by either all producers or only by low quality producers. If the highest threat of capture stems from offers that are accepted by all producers and the disclosure rule is noisy, credibility can be maintained by making honest certification more attractive to high quality producers. This in turn increases equilibrium investment levels as compared to full information disclosure.

The remainder of the chapter is organized as follows. Section 1.2 revies the related literature. The formal model is presented in section 1.3. Section 1.4 analyzes the static game in the absence of bribery. In section 1.5, we treat the general case of certification under the threat of capture. Section 1.6 concludes. All proofs are presented in the appendix.

[5]We analyze a belief system that substantially restricts the set of feasible disclosure rules. For different belief systems and sufficiently low discount factors, other (opaque) rules may be chosen by the certifier. The effect on social welfare is therefore not a general result.

1.2 Related literature

A stream of literature seeks to explain why certifiers often choose to only partially reveal quality information. Lizzeri (1999) finds that it is optimal for a monopolistic certifier in a static adverse selection environment to reveal almost no information. In this setting, this result is robust to introducing capture because a no revelation policy simply annihilates producer incentives to bribe. In the presence of moral hazard however, information revelation is necessary to create incentives for the provision of quality.

Albano and Lizzeri (2001) study optimal disclosure rules in a static model of both moral hazard and adverse selection. In their setting, a certifier chooses to employ noisy disclosure if his set of actions is restricted to flat fees. Kartasheva and Yilmaz (2013) explain imprecise ratings in a model with partially informed investors and heterogeneous liquidity needs of issuers. No disclosure is not optimal, because it deters high quality issuers from participating. With full disclosure, the fee is determined by the willingness to pay of the lowest certified type. Awarding to this type the best certificate with small but positive probability therefore allows for increasing the fee and also profits.

All papers mentioned above have profit maximization as the sole objective of the certifier. In contrast, this chapter suggests reputational concerns as another origin of a certifiers preference for coarse quality disclosure. In particular, the model we provide features full disclosure as a profit maximizing disclosure rule and nevertheless the certifier resorts to opacity because otherwise she cannot signal honesty.

Farhi, Lerner, and Tirole (2013) apply the term opacity to the disclosure of rejected applications for a certificate. In their model, a seller can turn to various certifiers, which differ in their acceptance of quality and whether they disclose true quality or only whether quality is not the lowest. The competition of certifiers makes 'rejection' a valuable information, whereas in our model there is only one chance to get certified and this information is worthless. Again, the certifiers preference for opacity stems from profit concerns. In a similar vein, Faure-Grimaud, Peyrache, and Quesada (2009) consider a model where the contract between the seller and the certifier entails the ownership of the rating, i.e. whether the seller can conceal it or not. If firms have only an imprecise signal of their own quality and some do not ask for a rating, simple ownership contracts emerge where the certification result is not published by the certifier but owned by the seller who can publish or conceal the result. In both studies the certifier's disclosure rule in terms of intrinsic product quality is exogenously given - opacity is referred to as the potential concealment of certificates and certification procedures and arises from profit concerns of the certifiers.

In Pagano and Volpin (2012) it is the seller who decides to release coarse information. In their model of rating asset-backed securities, the rating agencies' role is to confirm the information the issuer wants to conceal. It is again a profit concern that leads to opacity.

The threat of capture in certification markets has been analyzed by Strausz (2005). In a pure adverse selection setting with mandatory full disclosure, he analyzes the effects of a threat of capture on certification prices. He finds that in order to maintain credibility, for low discount factors, a certifier raises fees above the static monopoly price. This result is consistent with our finding: A larger fee reduces the share of certifying types and thereby increases the value of an uncertified product. As it is also the case in this chapter, it turns out that a major determinant for the certifiers credibility is the maximal bribe, which corresponds to largest difference in certification outcomes, e.g. the difference in values for the best certificate and uncertified products. A larger fee increases this cut-off but this implies that less information is revealed in equilibrium. Although this effect is also present in Strausz (2005), he however does not explicitly point it out. Credibility is maintained by reducing the maximal willingness to bribe. In Strausz (2005), this is affected by the value of not being certified, which, in turn, is an increasing function of the certification fee.

There is a rich literature on reputation building in markets with informational asymmetries. For example, Shapiro (1983) analyzes the forces at work when sellers build reputation. Biglaiser (1993) investigates the role of market intermediaries when sellers are unable to build their own reputation. Examples of works that treat reputational concerns of rating agencies are Mathis, McAndrews, and Rochet (2009) and Bolton, Freixas, and Shapiro (2012). In contrast to the chapter, these works follow the asymmetric information approach to reputations, where certifiers are assumed to always be committed (i.e. honest) with positive probability.[6] This, however, does not allow for studying the interaction between repuational concerns and information disclosure. The reason is that only false certification within a given disclosure rule can be studied, because with the committed certifier type the disclosure rule is already fixed and any departure reveals the certifier being not honest. Instead of assuming that

[6]See Mailath and Samuelson (2012, Chapter IV) for a discussion of this approach.

testing by the certifier is imperfect as is done in those works, we show how imperfect testing may endogenously arise in equilibrium.

We conclude this literature review by listing two related papers that do not focus on certification in particular. Levin (2003) extends the standard moral hazard setting to situations where contractual agreements are enforceable only to a certain degree and where reciprocal relations are long-term. The optimal contract derived by Levin has a coarse structure, which parallels our finding of coarse disclosure being optimal. As in this chapter, coarseness stems from a binding reputational constraint. The interaction of disclosure and incentives to exert effort is studied in Dubey and Geanakoplos (2010). In a model where a teacher seeks to induce effort by her students, it is shown that coarse grading schemes can help to induce all students to employ effort if they are disparate and care about their status in class.

1.3 Setup

We consider a dynamic framework in discrete time. A short-lived monopolistic producer is born in each period $t = 1, 2, \ldots, \infty$. He produces a single unit of quality $q_t \in \{q^l, q^h\}$, where $0 \leq q^l < q^h$. In the following, we refer to a *high type* producer if his product quality is q^h and to a *low type* producer otherwise. Prior to production, a producer chooses some investment level $e_t \in [0, 1]$. Quality is stochastic and the probability that the good is of high quality q^h is given by $Prob(q_t = q^h|e_t) = e_t$. This probability function is independent of t, i.e. quality levels are independent across time. Investment costs are given by the function $k(\cdot)$, which we assume is strictly increasing and strictly convex. For technical reasons we assume a non-negative third derivative, so that the certifier's profit function is concave and to guarantee interior solutions we additionally assume $k'(0) \leq q^l$ and $k'(1) \geq q^h$.

Consumers' reservation prices equal (expected) qualities. Both investment and quality level are private information to the producer. Consumers observe the product quality only after consumption. All other components of the model are common knowledge. A producer enters the market, decides upon the investment level e_t and the good is produced. At the end of each period goods are sold in a second-price auction[7] after which the producer leaves the market. Figure 1.1 summarizes the timing

[7] The second price auction results in a standard monopoly price that equals consumers' valuations. It circumvents signaling issues, e.g. letting the informed party take a publicly observed action that might be interpreted as a signal.

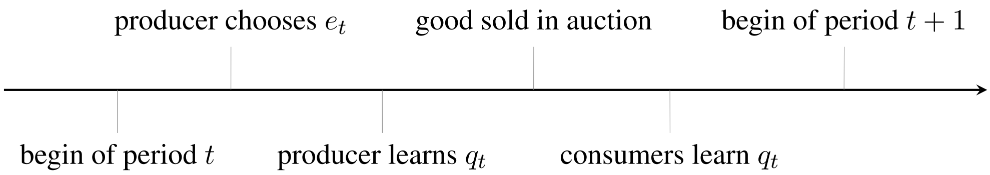

Figure 1.1: Timing in one period without certification

in period t. The equilibrium concept we use throughout the chapter is that of perfect Bayesian equilibrium.

To simplify notation, we set $q^l \equiv 0$ and define $v := q^h - q^l$. In the benchmark of *complete information* high quality goods are sold in the second-price auction at price v and low quality goods are sold at price 0. The producer then chooses e to maximize expected profits $ev - k(e)$. The first-best investment level e^* is thus given by $k'(e^*) = v$, which lies in the interval $[0, 1]$ due to our assumption $k'(1) \geq v$. In particular we have $e^* > 0$.

Under *asymmetric* information and in the absence of any further economic institution, a producer cannot persuade consumers that he offers a high quality good and the market price can therefore not be made contingent on a good's quality. It is standard to show that the Perfect Bayesian market outcome involves a market breakdown. In such an outcome, consumers form a belief $\widetilde{q}_t$ about the offered quality, which reflects their willingness to pay. In equilibrium, this belief has to be consistent with the actual expected quality, $E(q_t|e_t)$. Given any belief, the producer's optimal choice of investment is $e_t = 0$, as he maximizes $\widetilde{q}_t - k(e_t)$. Because $E(q_t|0) = 0$, in the unique equilibrium producers choose $e_t = 0$ and the quality of the good is zero in each period. The result is a market failure: high quality is never offered in equilibrium. We summarize this finding in the following lemma.

Lemma 1.1. *Without certification, producers choose $e_t = 0$ in each period. In equilibrium, quality is given by $q_t = 0$ and the price is 0 in each period.*

This inefficiency calls for the emergence of alternative market institutions to facilitate supply of high quality. The focus of the chapter lies on certification as one such institution. Assume that an infinitely long-lived certifier enters the market. She offers to disclose the result of some potentially imperfect test of the good's quality, prior to the good being sold. More precisely, at the beginning of the game, in period $t = 0$, the certifier announces a fee $f \geq 0$,[8] [9] and a disclosure rule $D = (\mathcal{C}, \alpha^l, \alpha^h)$.

Any producer who demands certification has to pay the fee f. The disclosure rule consists of a set $\mathcal{C} = \{C^1, \ldots, C^m\}$ of potential certificates and probability vectors α^l and α^h, where the k-th entry of vector α^i reflects the probability that a product of quality q^i is awarded certificate C^k whenever tested. We do not assume that these probabilities add up to one, i.e. we allow for $\sum_{k=1}^{m} \alpha_k^i < 1$. Hence, a product may remain uncertified with the conforming probability and will be sold as such. We assume that consumers cannot observe whether a product was tested, unless it is offered with a certificate.[10] Possible disclosure rules encompass for example *full disclosure*, where $\mathcal{C} = \{C^1, C^2\}$ and $\alpha^h = (0, 1)$ as well as $\alpha^l = (1, 0)$, or *no disclosure*, where $\mathcal{C} = \{C\}$ and $\alpha^i = (1)$.[11]

For a given certificate C^k, consumers form a belief $\widetilde{q}^{C^k}$ about the true quality of a product. The belief for uncertified products is denoted $\widetilde{q}^{\emptyset}$. For notational convenience we henceforth add $\emptyset$ to the set of certificates $\mathcal{C}$, which refers to uncertified products. Hence, $\mathcal{C} = \{C^1, \ldots, C^m, \emptyset\}$.

An interpretation of the disclosure rule, which we shall use throughout the chapter, is the following: the certifier can create any test that leads to a grading scheme with grades from the set $\mathcal{C}$ and results in the respective grades with conforming probabilities. This may be done with a computer program or a statistical test. In particular, after the test result is obtained, the certifier and the consumers share the same beliefs on product quality.

Finally we assume that the certifier's inspection costs are zero[12] and that she discounts future profits at rate $\delta \in (0, 1)$. Figure 1.2 illustrates the timing of the game

[8] Assuming a single fee f, that does not depend on the certificate, is without loss in the setting with only two quality levels. The best a certifier could do is, following the revelation principle, offering a menu of 'contracts' for the two potential producer types. Eventually, there is one payment referring to the high type and one referring to the low type. It can be easily shown that the optimal contract corresponds to the full disclosure rule, where high types pay f and low types pay 0 and true quality is revealed.

[9] The fee f creates a distortion as will become clear later on. The certifier could implement the first-best outcome, but only when moving first, i.e. when demanding an upfront payment *before* producers choose their investment. This timing however seems unreasonable in many certification markets.

[10] Hence products which "failed" the test are sold under the same label as products that didn't even take the test. This assumption is not crucial, since the certifier can replicate any outcome of a game where consumers are able to observe whether a product applied for certification.

[11] Note that certificates do not carry an intrinsic value. In the case that quality is fully revealed, whether C^1 or C^2 is the valuable certificate depends on the choice of α.

[12] This assumption simplifies the analysis without substantially affecting the results, which continue to hold as presented here for small but strictly positive inspection costs. Large inspection costs leave most of our results still valid, but create cumbersome case distinctions.

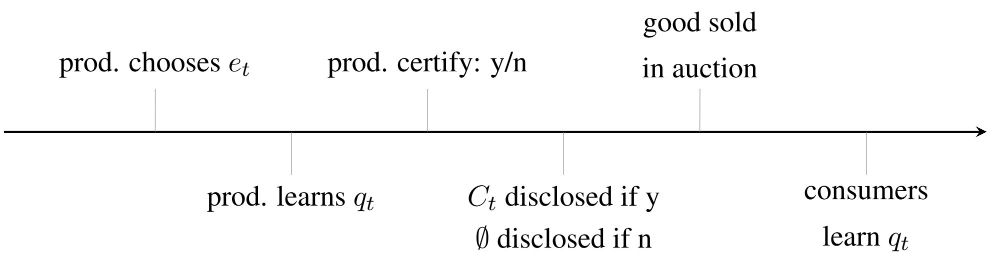

Figure 1.2: Timing of a period t with certification

with certification.

1.4 Optimal honest certification

In this section, we analyze certifier equilibrium strategies when the certifier is honest. By the stationary structure of the model, we can restrict our analysis to the certifier decision (D, f) plus a single period of production. Let $\pi^D(f)$ denote the equilibrium profit of the certifier, when adopting disclosure rule D with certification fee f.

We first study the case of full disclosure in some detail, as it will turn out that this disclosure rule can be used to achieve maximal profits. Consider the case that quality is fully revealed such that $\alpha^h = (1,0)$ and $\alpha^l = (0,1)$. Any product that is awarded C^1 is sold at a price v, whereas C^2 is worth nothing. The only plausible equilibrium is one where only high types apply for certification.[13] A producer chooses his investment according to

$$e = arg \max_{\widetilde{e}} \ \widetilde{e} \cdot (v - f) - k(\widetilde{e}). \tag{1.1}$$

This implies $k'(e) = v - f$ in equilibrium and certifier expected equilibrium profits can be expressed as

$$\widehat{\pi}^{FD}(e) = e \cdot (v - k'(e)). \tag{1.2}$$

Denote e^{FD} the equilibrium effort level under a full disclosure rule and f^{FD} the respective fee that maximizes certifier profits under full disclosure. The following

[13]Trivially, low quality producers do not demand certification when $f > 0$ since their revenues are most zero.

lemma proves that these values do exist and are unique.

Lemma 1.2. *Under full disclosure, there exists a unique fee f^{FD} that maximizes certifier profits. The uniquely defined equilibrium investment level e^{FD} is implicitly given by*

$$k''(e^{FD}) \cdot e^{FD} = v - k'(e^{FD}). \tag{1.3}$$

The fee is $f^{FD} = v - k'(e^{FD})$ and the (subgame-) equilibrium profit is $\pi^{FD} = e^{FD} \cdot f^{FD}$.

We continue analyzing general disclosure rules. The entire set of disclosure rules is complex and difficult to handle analytically. A closer look at equation (1.2), which allows us to express the certifier profit as function of the implemented investment level e, points to the advantages of using an indirect approach. We can express the attained profit of any certifier policy (D, f) solely in terms of the induced investment level e. This allows for a straightforward comparison of attained profits and leads us to the following proposition.

Proposition 1.1. *For any disclosure rule $D = \left(\mathcal{C}, \alpha^l, \alpha^h\right)$ and any fee $f \geq 0$, it holds that $\pi^D(f) \leq \pi^{FD}$ in equilibrium.*

Proposition 1.1 states that the certifier will always find it optimal to employ a full disclosure rule. The reason is that, investment incentives depend on the difference between payoffs from selling high and low quality products. Given full disclosure, the certification fee is sufficient to fully control this difference.

We conclude this section by pointing out that full disclosure is not the unique disclosure rule that yields the maximal certifier profit π^{FD}. First of all, one can implement the outcome of full disclosure with various disclosure rules by adding redundant certificates - either additional certificates for high types, which then all have the same value in equilibrium, or by adding certificates for low types that will not be issued in equilibrium. Because we assumed certification to be costless for the certifier, other rules equally achieve the maximal per-period profit: Issue two different certificates C^1 and C^2. Low quality products are only eligible for certificate C^2, hence $\alpha^l = (0, 1)$. High quality products receive certificate C^1 with probability $\alpha \in (0, 1)$ and C^2 otherwise, therefore $\alpha^h = (\alpha, 1 - \alpha)$. With this structure, it is possible to sustain an equilibrium in which all producer types demand certification.[14] The optimal certifier

[14]For this, we have to set the off-equilibrium belief $\widetilde{q}^{\emptyset} = 0$ and all other beliefs underly Bayesian updating.

profit π^{FD} is then obtained by choosing f and α appropriately.[15]

Disclosure rules of the latter kind play a crucial rule for the remainder of the chapter. We henceforth refer to them as *partial disclosure* rules.

1.5 The capture problem

So far we assume that the certifier sticks to the announced disclosure rule, in particular that she conducts the lottery honestly and grants the respective certificate. However, there is pressure from producers who wish to be awarded better certificates. For instance, if disclosure is meant to be noisy, a certifier might be willing to guarantee a producer a high value certificate in exchange for a bribe. In this section we address this issue by formally introducing the possibility of capture.

We follow Strausz (2005) in modeling the possibility of capture, using the framework of enforceable capture as initiated by Tirole (1986). Hence we assume that the certifier and the producer can write an enforceable side-contract with transfers. Consumers are fully aware of the possibility of these side-contracts, but cannot observe them.

Specifically, we model capture as follows: after a producer has learned his type q_t, but before deciding upon certification, the certifier, without observing q_t, may make an offer (C, b) to the producer. The offer consists of a certificate C, issued in case of acceptance, and a financial transfer b to be paid by the producer. The certifier thus offers to "sell" the sure certificate C at the price b, circumventing the customary certification procedure given by the disclosure rule. Hence, (C, b) are the terms at which she is willing to become captured. A producer, however, can reject this offer and, if willing to do so, insist on honest certification by paying the fee f. This last assumption is motivated following Kofman and Lawarrée (1993) in assuming that the certifier cannot forge certification without the help of the producer. Figure 1.3 displays the timing in a representative period t, allowing for the possibility of capture.

Note that the choice of the disclosure rule puts some limits on the set of *feasible* capture offers. For a general disclosure rule $D = \{\mathcal{C}, \alpha\}$ only offers of the form (C, b) with $C \in \mathcal{C}$ are feasible.[16]

[15] We formally show this in the proof of Proposition 1.6.

[16] This will be made more precise when formally introducing consumer beliefs. Granting a certificate which is not contained in D is certainly perceived as cheating by consumers. Consequently consumers believe to be faced with a worthless product and they will lose trust in the certifier's credibility.

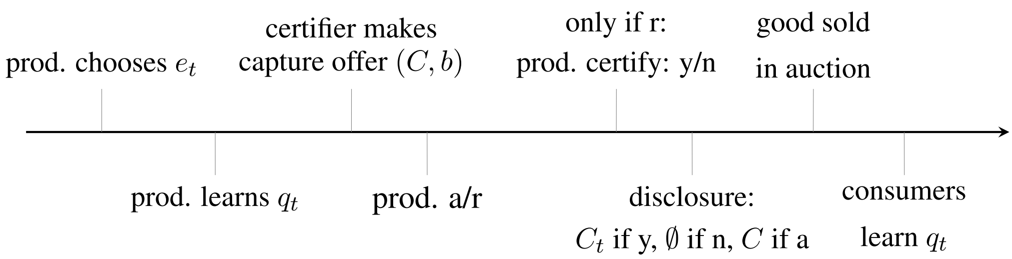

Figure 1.3: Timing of a period t with certification and capture

Within the framework presented here, capture may subvert honest certification for two reasons.[17] First, producers with low quality products are willing to side-contract with the certifier in order to obtain better certification. Second, high types may want to avoid uncertainty if disclosure is noisy.

In this section we are interested in the existence and characterization of equilibria where the certifier resists the temptation of making any capture offer of the above described kind. Throughout, we will work with different specifications of trigger beliefs. This becomes necessary as the ability of consumers to detect capture varies across disclosure rules. We assume consumers are able to perfectly observe quality after consumption. Therefore, if D is full disclosure or if certain certificates are awarded exclusively to high types, capture detection is also perfect.

Our particular idea behind the consumers' beliefs is the following: They stop trusting the certifier immediately if a false testimony about a product's quality is detected. Then, producers are not willing to pay for certification anymore. Consequently the certifier will lose future demand and makes zero profits henceforth. This prevents the certifier from becoming captured in the first place. We shall make this more precise in the following subsections.

1.5.1 Capture under full disclosure

Consider again the full disclosure rule introduced in section 1.4, i.e. there are the certificates C^1 and C^2 where C^1 is only awarded to high quality products. Because, by Proposition 1.1, a certifier would want to employ full disclosure whenever possible, we start by investigating capture under a full disclosure rule. We assume that

[17] When certification is costly for the certifier, there is a third reason: saving certification costs. As already mentioned in Footnote 12 our analysis can be extended to $c > 0$, but this involves some troubling case-by-case distinctions.

consumers trust certificates as long as they have not detected a deviation. A certifier who anticipates this behavior may be prevented from succumbing to the temptation of becoming captured by the fact that losing credibility will leave her without demand in future periods.

Denote $h_t = (C_t, q_t)$ the certification outcome in period t, where C_t is the issued certificate in period t and q_t is the true quality observed after consumption. If certification in period t did not take place, then $C_t = \emptyset$. Now let $H_t = (h_1, \dots, h_{t-1})$ summarize the history of certification at the beginning of period t. Finally, we denote $\widetilde{q}_t(C_t, H_t)$ a consumer's belief in period t when faced with a product carrying certificate C_t and when having observed history H_t. The following assumption on consumers' beliefs covers the described behavior.[18]

Assumption 1.1. *The consumers' beliefs* $\widetilde{q}_t(C_t, H_t)$ *satisfy* $\widetilde{q}_t(C_t, H_t) = \widetilde{q}^{C_t}$ *whenever* $\{\tau < t | q^{C_\tau} \neq q_\tau \vee C_\tau \notin \mathcal{C} \cup \{\emptyset\}\} = \emptyset$. *Moreover* $\widetilde{q}_t(C_t, H_t) = 0$ *whenever* $\{\tau < t | \widetilde{q}^{C_\tau} \neq q_\tau \ \vee \ C_\tau \notin \mathcal{C} \cup \{\emptyset\}\} \neq \emptyset$ *and* $\widetilde{q}_t(C_t, H_t) = 0$ *whenever* $C_t \notin \mathcal{C}$.

The assumption states that consumers trust the certifier if capture was not observed in the past. They however lose trust forever, once they detected cheating. Losing trust implies that consumers believe for any certifier's claim that the offered quality is zero.

With full disclosure, there are (at most) two types of bribing offers that can be made: (C^1, b) and (C^2, b). Obviously, an offer (C^2, b) is turned down by all types of producers, as it is worth nothing. Hence, in the following we focus on offers (C^1, b) and talk of a bribe b rather than (C^1, b). An offer b is accepted by high producer types whenever $b < f$. Low quality producers accept any bribe $b < v$ because acceptance will yield positive profits compared to zero profits for rejection.

In equilibrium, the certifier assigns probability $e(f)$ to the event that a producer is of high type, where $e(f)$ is the producer's optimal investment under full disclosure, derived from (1.1). We are interested in equilibria where capture does not occur. In all such equilibria, a producer chooses his optimal investment level knowing that he will not receive an acceptable capture offer. The acceptance probability $p(b|f)$ of bribing

[18] Note that consumers do not lose trust in the certifier when a product is awarded certificate C^2, although this should not happen in equilibrium. It is not necessary to include this case into consumers' beliefs, because any such event can only follow a non-profitable deviation by the certifier.

offer b given the certification fee f is given by

$$p(b|f) = \begin{cases} 1, & \text{if } b < f, \\ 1 - e(f), & \text{if } f \leq b < v, \\ 0, & \text{if } b \geq v. \end{cases} \tag{1.4}$$

We denote by $\Pi^D(f) = \sum_{t=1}^{\infty} \delta^{t-1}\pi^D(f) = \pi^D(f)/(1-\delta)$ the certifier's expected profit from honest certification under disclosure rule D and fee f. The certifier's expected profit from offering bribe b is denoted by $\widehat{\Pi}^D(b|f)$ and depends on whether the consumer detected capture as follows: whenever $b < f$, all producer types will accept the bribe, but only for low quality producers this is detected. Hence, $\widehat{\Pi}^{FD}(b|f) = b + e(f)\delta\Pi^{FD}(f)$. For $f \leq b < v$, only low quality producers accept the bribe and $\widehat{\Pi}^{FD}(b|f) = (1-e(f))b + e(f)(f + \delta\Pi^{FD}(f))$. Whenever $b \geq v$, all producers reject the bribe and the certifier obtains $\widehat{\Pi}^{FD}(b|f) = \Pi^{FD}(f)$.

If $\widehat{\Pi}^{FD}(b|f)$ exceeds $\Pi^{FD}(f)$ for some b, the certifier is actually better off becoming captured with the associated probability $p(b|f)$. We say that certification at a fee f is *capture proof* if and only if

$$\Pi^{FD}(f) \geq \widehat{\Pi}^{FD}(b|f) \tag{1.5}$$

for all b.

Note that $\widehat{\Pi}^{FD}(b|f)$ is increasing in b, both on $[0, f)$ and $[f, v)$ and it is constant for $b \geq v$. Furthermore $\widehat{\Pi}^{FD}(\cdot|f)$ is continuous at $b = f$.[19] Therefore, certifier profits from bribery are largest when b approaches v. Evaluating this yields the following proposition:

Proposition 1.2. *Under a full disclosure rule, an equilibrium satisfying Assumption 1.1 is capture proof. It exists if and only if*

$$\delta \geq \delta^{FD}(f) \equiv \frac{v}{v + \pi^{FD}(f)} \tag{1.6}$$

The proposition highlights the crucial role the discount factor plays for the existence of honest, i.e. capture proof, equilibria: the critical discount factor determines the relative weights of the short run gain - the bribe b - and the long run loss of capture

[19] To see this compare the left and right limit: $\lim_{b \downarrow f} \widehat{\Pi}^{FD}(b|f) = f + e(f)\delta\Pi^{FD}(f) = (1 - e(f))f + e(f)\big(f + \delta\Pi^{FD}(f)\big) = \lim_{b \uparrow f} \widehat{\Pi}^{FD}(b|f)$.

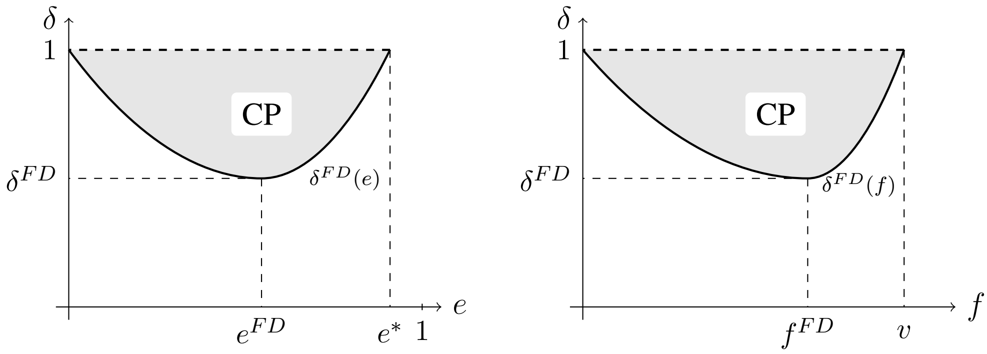

Figure 1.4: Capture proof combinations of (e, δ) resp. (f, δ) under full disclosure.

- foregone future profits from certification. To see this, note that all bribes $b < v$ are accepted with some positive probability and therefore, the largest possible short-run gain equals v. In the long run, a certifier risks her per-period profits $\pi^{FD}(f)$.

Because the certification fee enters (1.6) only via the per-period profit, $\delta^{FD}(f)$ depends on f only through $\pi^{FD}(f)$, which is concave in f. Therefore $\delta^{FD}(f)$ must be convex in f and minimized at the profit maximizing fee f^{FD}.

Corollary 1.1. *For any discount factor $\delta \geq \delta^{FD}$ there exists an interval of fees $[f_l(\delta), f_h(\delta)]$, which sustains capture-proof certification under full disclosure, where*

$$\delta^{FD} \equiv \frac{v}{v + \pi^{FD}}. \tag{1.7}$$

As an immediate consequence we get that that the static monopoly fee f^{FD} can sustain honest certification for all discount factors $\delta \geq \delta^{FD}$. The right panel of Figure 1.4 depicts the set of feasible (δ, f)-combinations for full disclosure.

Alternatively one might ask the question, what level of producer investment can be implemented via capture-proof certification with a full disclosure rule? The analysis follows the same arguments as above, only that certifier profits in the inequality of Proposition 1.2 are expressed in terms of e.

Proposition 1.3. *For any $\delta \geq \delta^{FD}$ there exist values $e_l^{FD}(\delta) < e_h^{FD}(\delta)$ such that an investment level e can be implemented in a capture-proof equilibrium if and only if $e \in [e_l^{FD}(\delta), e_h^{FD}(\delta)]$. A particular investment level $e \in [0, e^*]$ can be implemented in a capture-proof equilibrium with full disclosure if and only if*

$$\delta \geq \delta^{FD}(e) \equiv \frac{v}{v + e \cdot (v - k'(e))}. \tag{1.8}$$

The set of feasible (e, δ)-combinations is depicted in the left panel of Figure 1.4. Note that the first-best investment level e^* can only be (virtually) implemented for $\delta = 1$. Whenever $\delta < 1$, fees must be strictly positive in order to induce the certifier to remain honest. But then, the producer does not obtain the entire return on his investment. Hence, it must be that $e < e^*$.

1.5.2 Capture under partial disclosure

We next argue that alternative noisy disclosure rules can improve certifier credibility in the sense that they increase the range of discount factors that allow for capture-proof equilibria.

To gain some intuition consider again condition (1.6). This condition summarizes the trade-off between short-run gains and long-run losses. A larger profit $\pi^D(f)$ reduces the critical discount factor and full disclosure guarantees maximal per-period profits. On the other hand, $\delta^{FD}(f)$ is decreasing in v, which represents the the maximal bribe still accepted by low-type producers and therefore the largest possible short-run gain from capture.

Using noisy disclosure the certifier can affect the maximal short-run gain in various dimensions. First of all, lowering the value of the best certificate or increasing the value of the worst certificate (resp. the value of uncertified products) decreases the gap between particular certification outcomes. This effect can be used to reduce the maximal bribe which producers are willing to pay. Second, with noisy disclosure the certifier can sustain an outcome where both producer types demand certification. Upon colluding with a producer type the certifier foregoes the regular certification fee, which reduces the effective gain from becoming captured.

Before analyzing noisy disclosure rules, we have to reconsider the detection possibilities by consumers. An implication of noisy rules is that consumers may hold probabilistic beliefs about a product's quality. In order to simplify matters and because it suffices to make our point clear, we focus on partial disclosure rules as introduced in section 1.4. Other noisy disclosure rules are discussed in section 1.6 and in the appendix.

Under partial disclosure, there are again two certificates C^1 and C^2, where certificate C^1 is awarded exclusively to high quality products and C^2 is awarded to a high quality seller with probability $1 - \alpha$ and to every low quality seller. With an appropriately chosen fee f all producer types demand certification, hence there are no uncertified products on the equilibrium path. The corresponding off-equilibirum belief is $q^{\emptyset} = 0$. The fact that C^1 is awarded exclusively to high quality products makes effective trigger punishment possible. In particular, it then suffices that the certifier is punished only if probability zero events (a low quality product was awarded certificate C^1) are observed. The fact that capture detection is not possible if bribes are being paid in exchange for the low value certificate C^2, which is assigned to both high and low types, turns out not to be crucial. This relies on the fact that in the equilibria under consideration all producer types demand certification, hence receiving certificate C^2 is the worst possible outcome. Certificate C^2 can therefore not be part of a profitable bribing offer, as we will argue later.

To specify consumer beliefs, let $h_t = (C_t, q_t)$ denote the certification outcome in period t and, as before, $H_t = (h_1, \ldots, h_{t-1})$ describes the history in period t. Consumer's beliefs are specified as follows

Assumption 1.2. *The consumers' beliefs $\widetilde{q}_t(C_t, H_t)$ satisfy $\widetilde{q}_t(C_t, H_t) = \widetilde{q}^{C_t}$ whenever $\{\tau < t | Prob(C = C_\tau | q = q_\tau) = 0 \ \vee \ C_\tau \notin \mathcal{C} \cup \{\emptyset\}\} = \emptyset$. Moreover $\widetilde{q}_t(C_t, H_t) = 0$ when either $C_t \notin \mathcal{C}$ or $\{\tau < t | Prob(C = C_t | q = q_t) = 0 \ \vee \ C_\tau \notin \mathcal{C} \cup \{\emptyset\}\} \neq \emptyset$.*

Note that in contrast to Assumption 1.1, the consumer trust the certifier unless probability zero events occured in the past. Because the crucial bribe entails certificate C^1, which is exclusively awarded to high quality producers, this essentially says that consumers stop trusting the certifier, whenever they find a low quality product carrying certificate C^1. Cheating on the lottery leading to certificate C^2 is not detected and also not punished, but because this certificate corresponds to the worst outcome this will not happen as a result of a capture offer.

Bribing offers can now be of two kinds: (C^1, b) and (C^2, b). Offer (C^2, b) is never beneficial. It would only be accepted for $b < f$, because any producer receives at least the certificate C^2 when applying for (honest) certification and the certifier gets f from any producer who is honestly tested. Thus, we can focus on bribing offers of the form (C^1, b), which we will simply refer to as b. Recall that certificate C^1 can only be awarded to high quality products. Hence, $q^{C^1} = v$. To simplify notation, denote V_2 the value of a C^2-certified product, i.e. $V_2 = q^{C^2}$. Furthermore, recall that α is the probability with which a high type is awarded C^1.

A bribe b is accepted by low types whenever $V_2 - f < v - b$. High quality producers accept b if $\alpha v + (1 - \alpha) V_2 - f < v - b$. Denote $e(\alpha)$ the equilibrium investment.[20]

[20]The investment decision does not depend on the fee because in equilibrium, all types apply for

Then bribery acceptance probabilities are

$$p(b|\alpha, f) = \begin{cases} 1, & \text{if } b < f + (1-\alpha)(v - V_2), \\ 1 - e(\alpha), & \text{if } f + (1-\alpha)(v - V_2) \le b < f + (v - V_2), \\ 0, & \text{if } b \ge f + (v - V_2). \end{cases}$$

Let $\Pi^{PD}(\alpha, f)$ denote the expected profit from applying a partial disclosure rule and honestly disclosing information in each period. The corresponding expected certifier profits from bribing offer b are

$$\widehat{\Pi}(b|\alpha, f) = \begin{cases} b + e(\alpha)\delta\Pi^{PD}(\alpha, f), & \text{if } b < f + (1-\alpha)(v - V_2), \\ (1 - e(\alpha))b & \text{if } f + (1-\alpha)(v - V_2) \\ \quad +e(\alpha)\left(f + \delta\Pi^{PD}(\alpha, f)\right), & \quad \le b < f + (v - V_2), \\ \Pi^{PD}(\alpha, f), & \text{if } b \ge f + (v - V_2). \end{cases}$$

Note that whenever high types accept the bribery offer, this is not perceived as cheating because the certificate then matches the observed quality level. The function $\widehat{\Pi}(b|\alpha, f)$ is increasing in the respective subintervals. But, contrary to the respective case of full disclosure, it exhibits a downward-jump at $b = f + (1-\alpha)(v - V_2)$. The reason is that high types are willing to accept bribes strictly larger than the certification fee f to avoid the lottery between the good and the bad certificate. Therefore, at least locally, the certifier is better off bribing all producers instead of only the low types as it was the case with full disclosure. Furthermore, the maximal bribe that is accepted by at least some types is now $f + v - V_2$, which is weakly lower than under full disclosure, where the maximal bribe is v.[21]

Revisiting condition (1.5), we say certification at fee f with noise level α is capture-proof, if and only if

$$\Pi^{PD}(\alpha, f) \ge \widehat{\Pi}(b|\alpha, f). \tag{1.9}$$

certification and therefore pay f anyway. The expected producer profit is $e(\alpha V_1 + (1-\alpha)V_2) + (1 - e)V_2 - f - k(e)$ and consequently the optimal investment level depends on α but not on f.

[21] In order to have all producer types demand certification it has to hold that $f \le V_2$. Consequently $f + v - V_2 \le v$.

Analyzing this latter condition yields the following proposition.

Proposition 1.4. *With partial disclosure, an equilibrium satisfying Assumption 1.2 is capture-proof. It exists if and only if*

$$\delta \geq \delta^{PD}(\alpha, f) \equiv \max\left\{\delta^{l}(\alpha, f), \delta^{l,h}(\alpha, f)\right\}, \tag{1.10}$$

where $\delta^{l}(\alpha, f) = \frac{v-V_2}{v-V_2+f}$ *and* $\delta^{l,h}(\alpha, f) = \frac{(1-\alpha)(v-V_2)}{(1-\alpha)(v-V_2)+(1-e(\alpha))f}$.

The result gives a lower bound on the discount factor δ to guarantee existence of a capture-proof equilibrium with partial disclosure. The critical discount factor discount factor $\delta^{PD}(\alpha, f)$ depends on the parameters in the way how they affect short-run gain and long-run loss from capture and on which producer types accept the bribing offer that yields largest deviation profits.

The term $\delta^{l}(\alpha, f)$ refers to the case where the largest threat stems from bribes accepted only by low types. The numerator $v - V_2$ is the effective bribe, defined as the bribery payment minus foregone payments. In the denominator we find again the effective bribe and the per-period profit f, reflecting the long-run loss from capture. The term $\delta^{l,h}(\alpha, f)$ refers to the case where the largest threat stems from bribes accepted by all types. Here the effective bribe is $(1-\alpha)(v-V_2)$. Because the long-run profit is only at stake if quality is low, long-run profits are lost with probability $(1-e(\alpha))$. Although the classical trade-off between short-run gain and long-run loss, that we already identified for full disclosure, prevails, the derivation of the maximal short-run gain is more involved for partial disclosure.

From Proposition 1.4 we identify a third notable difference between capture under full and noisy disclosure. Short-run gains from capture can be reduced due to the different equilibrium structure: all producers certify in equilibrium which implies that the certifier always loses fee payments if he is captured. Therefore, a larger fee f not only increases the long-run losses but at the same time reduces the short-run gains from capture.

It is now straightforward to see that $\delta^{PD}(\alpha, f)$ is decreasing in the certification fee f. This implies that for any partial disclosure rule (i.e. any α) the threat of capture is lowest when f is maximal. To keep all producers applying for certification, f cannot exceed V_2. It is therefore optimal to set $f = V_2$, which leaves low quality producers with an expected profit of zero. The following corollary summarizes.

Corollary 1.2. *With partial disclosure a capture-proof equilibrium satisfying Assumption 1.2 exists if and only if*

$$\delta \geq \delta^{PD}(\alpha) \equiv \max\left\{\delta^{l}(\alpha), \delta^{l,h}(\alpha)\right\}, \tag{1.11}$$

where $\delta^{l}(\alpha) = \frac{v-e(\alpha)(v-k'(e(\alpha)))}{v}$ *and* $\delta^{l,h}(\alpha) = \frac{1}{1+e(\alpha)}$.

Corollary 1.2 allows us to reduce the problem of finding the critical discount factor for partial disclosure to the one-dimensional problem of finding the optimal level of α, the probability that high quality is revealed. In fact, $\delta^{PD}(\alpha)$ depends on α only through the equilibrium value for producer investment $e(\alpha)$. The set of investment levels that can be implemented by partial disclosure is $(0, e^*)$, the same set as for full disclosure. Defining $\delta^{PD} \equiv \min_\alpha \delta^{PD}(\alpha)$ allows us to formulate the analog of Proposition 1.3 for partial disclosure.

Proposition 1.5. *For any* $\delta \geq \delta^{PD}$ *there exist values* $e_l^{PD}(\delta) < e_h^{PD}(\delta)$ *such that an investment level* e *can be implemented in a capture-proof equilibrium if and only if* $e \in [e_l^{PD}(\delta), e_h^{PD}(\delta)]$. *A particular investment level* $e \in [0, e^*]$ *can be implemented in a capture-proof equilibrium with noisy disclosure if and only if*

$$\delta \geq \delta^{PD}(e) = \max\left\{\delta^{PD,l}(e), \delta^{PD,l,h}(e)\right\} \tag{1.12}$$

where $\delta^{PD,l}(e) = \frac{v-e(v-k'(e))}{v}$ *and* $\delta^{PD,l,h}(e) = \frac{1}{1+e}$.

Proposition 1.5 makes implementation of capture-proof equilibrium under full and partial disclosure directly comparable. Before investigating this in the next section we want to highlight some properties of the function $\delta^{PD}(e)$. Writing $e(v - k'(e)) = \pi^{PD}(e) = f$ the term $\delta^{PD,l}(e)$ can be expressed as $(v - f)/(v - f + \pi^{PD}(e))$. This resembles the trade-off between short-run gain and long-run loss, already identified above. Only the maximal short-run gain with partial disclosure is the maximal bribe minus foregone regular payments. The same trade-off leads to $\delta^{PD,l,h}(e)$, which is, however, independent of the producer's cost function $k(e)$.

The maximal bribe that is accepted from both producer types in particular must be accepted from high quality producers. For them, the difference between the sure certificate C^1 and the lottery faced when certifying honestly matters. This difference is closely related to a producers' investment incentives, in fact one can show that the maximal bribe equals $v - k'(e)$. Both short-run gain and long-run loss depend in a

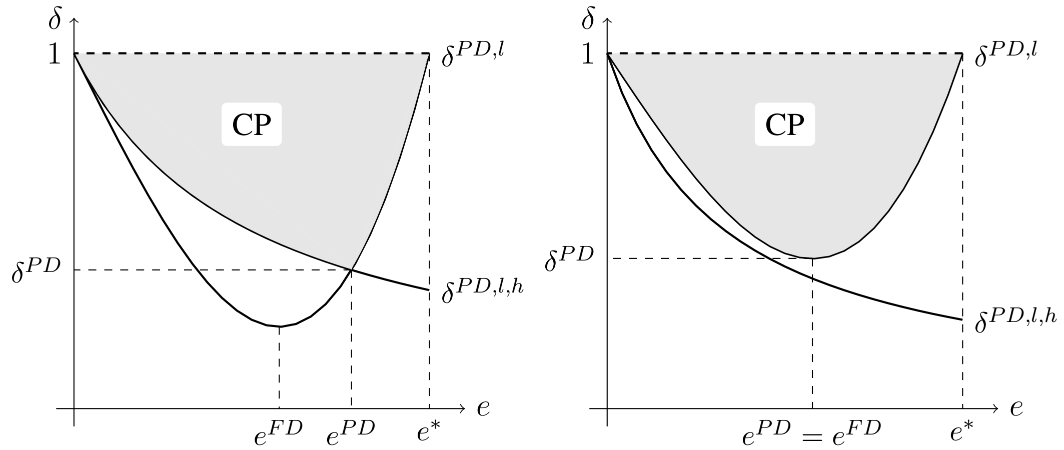

Figure 1.5: Capture-proof (e, δ)-combinations for low (left) and high (right) marginal costs k' at $e = e^{FD}$.

similar way on the investment incentives[22], consequently the fraction $\delta^{PD,l,h}(e)$ does not depend on the producers cost function anymore.

Which of the two terms, $\delta^{PD,l}(e)$ and $\delta^{PD,l,h}(e)$, is larger? $\delta^{PD,l,h}(e)$ is decreasing in e, starting at 1 for $e = 0$ towards $1/2$ for $e = 1$. On the other hand $\delta^{PD,l}(e)$ is convex in e with a unique minimum at $e = e^{FD}$. Furthermore $\delta^{PD,l}(0) = \delta^{PD,l}(1) = 1$. Therefore, δ^{PD} is either $\delta^{PD,l}(e^{FD})$, that is the minimum of $\delta^{PD,l}$, or it is the intersection of both fractions lying to the right of $e = e^{FD}$. Figure 1.5 illustrates the two cases, the latter in its left part.

1.5.3 Sub-optimality of full disclosure

In the previous sections, we identified the conditions under which capture-proof equilibria exist for full disclosure and a special class of noisy disclosure rules. These conditions are expressed in terms of the critical discount factors δ^{FD} and δ^{PD}. It is the aim of this chapter to show that opaque disclosure rules can be used by the certifier to improve his credibility. Comparing the critical discount factors δ^{FD} and δ^{PD} is short-hand for comparing the entire sets of (e, δ)-combinations, for which a capture-proof equilibrium exists with the respective disclosure rule.

We prove in this section that the two sets are different and, more importantly, that the respective set for full disclosure is contained in the respective set for partial

[22]As discussed, the short-run gain equals $v - k'(e)$. The long-run loss is the per-period profit, which was already shown to be $e(v - k'(e))$.

disclosure. Consequently there exists an intermediate range of discount factors for which there does not exist a capture-proof equilibrium with full disclosure, but it is still possible to sustain capture-proof equilibria with partial disclosure.

As we have discussed several times throughout the chapter, the key trade-off for implementing a capture-proof equilibrium is short-run gains versus long-run losses. Either disclosure rule leads to a per-period profit of $\pi(e) = e(v - k'(e))$ when implementing effort level e, the potential long-run loss is therefore the same. However, with partial disclosure the short-run gain from becoming captured by only low quality producers is $v - f$, compared to v for full disclosure. The resulting trade-off is resolved in favor of partial disclosure. So far this assumes that the largest threat of capture indeed stems from low quality producers. Although this is in general true for full disclosure, it ceases to hold for partial disclosure.

When the maximal threat stems from a bribe accepted by all producer types, the long-run loss is reduced. Only when the producer is of low quality this is perceived as cheating by consumers and punished accordingly. So per-period profits are only lost with probability $1 - e$. On the other hand such a bribe must be smaller in order to be acceptable for high quality producers, which reduces the short-run gain. The following proposition proves that the latter effect outweighs the former.

Proposition 1.6. *It holds that $\delta^{PD} < \delta^{FD}$. For any $\delta \in [\delta^{PD}, \delta^{FD}]$, a capture-proof equilibrium can only be sustained applying a noisy disclosure rule. Furthermore, for any $\delta \geq \delta^{FD}$, we have that $[e_l^{FD}(\delta), e_h^{FD}(\delta)] \subsetneq [e_l^{PD}(\delta), e_h^{PD}(\delta)]$.*

Proposition 1.6 shows our main result that opacity can be used as a tool to improve certifier credibility. For any level of producer investment e, the range of discount factors that allow for capture-proof implementation of e is strictly larger for partial disclosure compared to full disclosure. Similarly, for any discount factor δ, the set of investment levels that are implementable in a capture-proof equilibrium with partial disclosure is strictly larger then the corresponding set for full disclosure. The superiority of partial disclosure therefore goes along two dimensions. Figure 1.6 displays these differences. The dark-grey area corresponds to (e, δ)-combinations that can be implemented as a capture-proof equilibrium under full disclosure. The light-grey area shows the *additional* (e, δ)-pairs that allow for implementation in capture-proof equilibrium under partial disclosure.

In Section 1.4, we have argued that a certifier would always want to implement e^{FD} as this maximizes her per-period profits. With full disclosure, this is only possi-

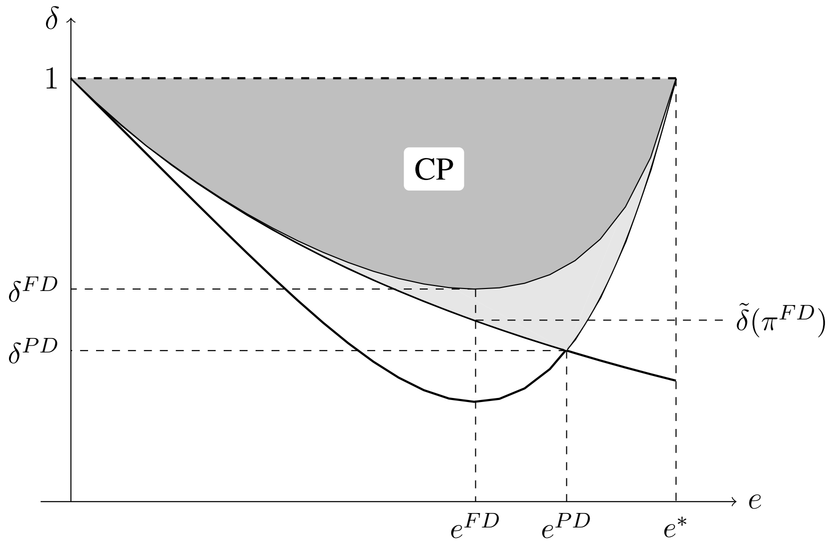

Figure 1.6: Dark-grey: capture-proof certification with full disclosure. Light-grey: (additional) capture-proof certification with noisy disclosure.

ble when $\delta \geq \delta^{FD}$. Partial disclosure allows for capture-proof equilibria also for lower discount factors. It is remarkable that, at least for a range of discount factors, this can be achieved without waiving any profits. To see this, denote $\tilde{\delta}(\pi^{FD})$ the smallest discount factor, such that a capture-proof equilibrium is sustained and achieves per-period profits of π^{FD}. The following corollary is an immediate consequence of Proposition 1.6.

Corollary 1.3. *It holds that*

$$\tilde{\delta}(\pi^{FD}) = \max\left\{\frac{v - \pi^{FD}}{v}, \frac{1}{1 + e^{FD}}\right\} < \delta^{FD}.$$

1.5.4 Welfare properties of partial disclosure

In this subsection, we study welfare properties of capture-proof equilibria with partial disclosure. When $\tilde{\delta}(\pi^{FD}) = \delta^{PD}$ we also have $\delta^{PD} = (v - \pi^{FD})/v$. In this case, the largest threat of capture stems from low quality producers, i.e. the largest deviation profit for the certifier is achieved for $b = v$. Then the certifier can still achieve the maximal per-period profits π^{FD} in a capture-proof equilibrium for any $\delta \geq \delta^{PD}$, which implies implementing $e = e^{FD}$.

This is however not true when $\tilde{\delta}(\pi^{FD}) > \delta^{PD}$. As can be seen from Figure 1.6, for discount factors below $\tilde{\delta}(\pi^{FD})$ the profit maximizing level of investment e^{FD} is no longer capture-proof implementable. Instead only larger values of producer investment can be implemented when $\delta \in [\delta^{PD}, \tilde{\delta}(\pi^{FD}))$. To provide an intuition for this, note the following: Bribing offers b that are accepted by all producer types pose the largest threat. Now, implementing a larger e leads to a reduction in V_2, as otherwise profits would increase beyond π^{FD}. To incentivize producers to make larger investments the certifier has to increase α.

As now shown, for high quality producers the difference in expected profits between the lottery of the certification process and the sure certificate v is reduced.[23] This in turn lowers the maximum bribe they are willing to pay for capture and therefore reduces the short-run gain for the certifier from any such offer. From a welfare perspective this increase in investment is beneficial. Social welfare is given by $e \cdot v - k(e)$ in each period. The first-best investment level e^* was shown to be strictly larger than e^{FD} and welfare is strictly increasing on $[0, e^*]$. Implementing certification with partial disclosure for discount factors $\delta \in [\delta^{PD}, \tilde{\delta}(\pi^{FD})]$ therefore increases social welfare compared to doing so for larger levels of the discount factor. Put differently, a severe threat of capture increases welfare. We summarize this in the following proposition.

Proposition 1.7. *Assume $\tilde{\delta}(\pi^{FD}) > \delta^{PD}$. For intermediate levels of the discount factor, i.e. $\delta \in [\delta^{PD}, \tilde{\delta}(\pi^{FD}))$, only investment levels that are strictly larger than e^{FD} can be capture-proof implemented with partial disclosure. This leads to increased social welfare.*

1.6 Discussion

We analyze the effects of reputational concerns on optimal disclosure rules from the point of view of a monopolistic certifier. Our main finding is that if capture is an issue, a certifier benefits from resorting to coarser certification in order to reduce the threat of capture and this is indepent of a potential profit concern that pushes the certifier in the same direction. In particular, for medium discount factors, sustaining honest certification is impossible if information is fully disclosed whereas it is still possible if information disclosure is noisy.

[23] Honest certification yields an expected payoff $\alpha v + (1 - \alpha)V_2$. This value is reduced when α increases and V_2 decreases at the same time.

Implications of our analysis are manifold. First of all we provide a novel explanation for the occurrence of imperfect testing. In many papers on e.g. rating agencies (examples include Mathis, McAndrews, and Rochet (2009) and Bolton, Freixas, and Shapiro (2012)) imperfect testing is exogenously given, whereas here it arises in equilibrium. An empirical implication is that for low discount factors we expect disclosure to be coarser. This is consistent with the casual observation that certification in markets with low volume, such as wine, technical inspections or eco-labels often involves only a few different certificates. On the other hand, the high volume rating market exhibits a rather wide variety of different but still coarse certificates per rating agency.

Our findings also have important policy implications. Politics tend to push certifiers to precisely reveal information.[24] Our results suggest that doing so may lead to unforeseen consequences for the functioning of those markets, as it might become more difficult to build up a reputation and resist capture if certificates are required to be too precise. Similarly, regarding the current financial crisis, forcing rating agencies to issue more precise information might even exacerbate capture problems.

We demonstrate our results in a highly stylized model, but the intuition behind our results is general. In particular, they carry over to more than only two quality specifications. This makes the analysis simpler on the one hand, as it can be shown that already coarse deterministic disclosure rules outperform full disclosure. On the other hand the analysis is complicated by the fact that full disclosure is not necessarily optimal anymore, when capture is ignored. The first point already becomes clear from a setting with three quality levels. Full disclosure can then entail both the highest and the medium quality producer demand certification. A coarse rule would specify one certificate awarded to all but low quality. Obviously, for both rules the same types of producer demand certification. In the latter case however the maximal bribe is strictly lower. For similar investment levels and fees, the critical discount factor is therefore strictly lower for the coarse rule. The precise analysis is more complicated, since the coarse rule generates different investment incentives for producers. In Appendix B we offer an illustration for a special case of probability distributions.

We point out that our restriction to a particular class of noisy disclosure rules is without loss of generality. First, offering various coarse certificates generates incentives for the certifier to always offer the best among the noisy certificates in a bribing offer. This will be accepted (at least by low quality producers) in order to avoid a

[24] Such as in the Dodd-Frank Act, see Footnote 4

lottery that includes the worst certificates. As deviations of this kind remain undetected they will occur with certainty, that destroys the equilibrium. Second, disclosure rules that do not allow for unambiguous detection of deviations call for a different type of trigger beliefs. Consumers lose trust in the certifier whenever they first detect low quality sold with the best certificate. This leads to punishments even if collusion did not take place. The harsher punishments makes it impossible to sustain capture proof equilibria for low discount factors. Proposition 1.8 in the appendix makes this statement precise.

Also the assumptions on the certifier's policy space is not restrictive. In our two period model, a full disclosure rule is equivalent to certificate-dependent payments: for a good certificate pay the certification fee f and for the bad certificate pay zero. By the revelation principle an optimal policy involves a certification contract for each of the two quality types. It is then straightforward to verify that the optimal policy is outcome-equivalent to a full disclosure rule with a flat fee.

Finally we use a specific extensive form to model capture. More sophisticated forms to study imply non-uniform bribing offers, e.g. menus, to elicit the producers' private information. Also, later bribing, after the certifier learned q or giving producers the possibility to signal their private information are possible extensions. The exact extensive form may well affect parts of the analysis, but the main finding of the advantage of opacity does not depend on the specific extensive form.

Appendix

1.A Proofs

1.A.1 Proofs of Section 1.3

Proof of Lemma 1.1. Follows immediately from the arguments given in the text. □

1.A.2 Proofs os Section 1.4

Proof of Lemma 1.2. Following the arguments given in the text the certifier maximizes (1.2). Recall that we assume $k'''(\cdot) \geq 0$, which ensures that this profit function is concave in e, thus the first-order condition is sufficient for an optimum. This first-order condition is $0 = v - k'(e) - ek''(e)$. Define $\Psi(e) = v - k'(e) - ek''(e)$. We have $\Psi(0) = v > 0$ and $\Psi(1) = v - k'(1) - k''(1) \leq 0$ by our assumptions on $k(\cdot)$. Furthermore Ψ is strictly decreasing due to strict concavity of $k(\cdot)$. Hence there exists a unique e^{FD} such that $\Psi(e^{FD}) = 0$, which consequently is the unique maximizer of the certifier profit. The formulas for e^{FD} and f^{FD} follow easily from the formulas above. □

Proof of Proposition 1.1. First of all a disclosure rule can potentially lead to four different subgames: (1) no producer demands certification, (2) only low quality producers demand certification, (3) only high quality producers demand certification, and (4) all producers demand certification. Note that we do not explicitly consider mixed strategies by producers. The reason is that any outcome where some producers randomize their certification decision can be replicated by a disclosure rule that adds the respective probabilities for not certifying to the probabilities of remaining uncertified though paying for certification. To see this, assume type i chooses to certify with probability $\gamma \in (0,1)$. Now multiply every α^i by γ and increase the probability of

remaining uncertified appropriately. After changing the fee from f to γf, it is easy to see that this adjusted disclosure with the reduced fee leads to the same investment incentives and also to the same equilibrium prices for (un-)certified products and the certifiers profit is unchanged.
Case (1) trivially leads to zero profits and the claim is proven.
Case (2) leads to consumers paying zero in equilibrium for certified products.[25] To make low quality producers "pay" for certification we consequently must have $f = 0$ which leads to zero profits and proves our claim also in this case.
Case (3) can be analyzed as follows: If only high types certify, rational behavior by consumers dictates that a certified product is sold at a price v. Uncertified products however can be of either high or low quality and have some price $q^{\emptyset} \in [0, v)$.
A producer's investment decision is given by the solution of

$$\max_{e} \; e\Big(\sum_k \alpha_k^1 v + (1 - \sum_k \alpha_k^1) q^{\emptyset} - f\Big) + (1-e)q^{\emptyset} - k(e),$$

which yields the following first-order condition for producer investment:

$$\Big(\sum_k \alpha_k^1 (v - q^{\emptyset}) - f\Big) = k'(e).$$

Rewriting this constraint in terms of induced investment yields $f = v - k'(e) - (1 - \sum_k \alpha_k^1)(v - q^{\emptyset}) - q^{\emptyset}$. Now we have for the certifier profit

$$\begin{aligned}\pi^D(f) &= e(f, D) \cdot f \\ &= e \cdot \big(v - k'(e) - (1 - \sum_k \alpha_k^1)(v - q^{\emptyset}) - q^{\emptyset}\big) \\ &\leq e \cdot (v - k'(e)) \leq \pi^{FD}.\end{aligned}$$

This proves the claim for case (3).
Finally consider case (4): When both producer types demand certification, the resulting certifier profit in the subgame is $\pi^D(f) = f$. The price at which a product holding certificate C^i can be sold is

$$q^{C^i} = v \cdot \frac{e\alpha_i^h}{e\alpha_i^h + (1-e)\alpha_i^l}.$$

[25] A disclosure leading to this particular subgame is given by $\mathcal{C} = \{C\}$, $\alpha^l = 1$ and $\alpha^h = 0$.

Uncertified products are sold at price $q^{\emptyset} = v \cdot \frac{e\left(1-\sum_i \alpha_i^h\right)}{e\left(1-\sum_i \alpha_i^h\right)+(1-e)\left(1-\sum_i \alpha_i^l\right)}$. A producer's investment decision follows from maximizing his expected payoff from certification, given by

$$e \cdot \left(\sum_i \alpha_i^h q^{C^i} + \left(1 - \sum_i \alpha_i^h\right) q^{\emptyset}\right) + (1-e) \cdot \left(\sum_i \alpha_i^l q^{C^i} + \left(1 - \sum_i \alpha_i^l\right) q^{\emptyset}\right) - f - k(e).$$

The resulting investment constraint is

$$k'(e) = \sum_i (\alpha_i^h - \alpha_i^l)(q^{C^i} - q^{\emptyset}). \tag{1.13}$$

On the other hand, from the formula given for q^{C^i} we have $e\alpha_i^h q^{C^i} + (1-e)\alpha_i^l q^{C^i} = ev\alpha_i^h$. Similarly $e(1-\sum_i \alpha_i^h)q^{\emptyset} + (1-e)(1-\sum_i \alpha_i^l)q^{\emptyset} = ev(1-\sum_i \alpha_i^h)$. Summing those expressions yields

$$\sum_i \left(e\alpha_i^h q^{C^i} + (1-e)\alpha_i^l q^{C^i}\right) + e(1-\sum_i \alpha_i^h)q^{\emptyset} + (1-e)(1-\sum_i \alpha_i^l)q^{\emptyset} = ev. \tag{1.14}$$

Rewriting the left hand side of equation (1.14) yields

$$e\sum_i (\alpha_i^h - \alpha_i^l)(q^{C^i} - q^{\emptyset}) + \sum_i \alpha_i^l q^{C^i} + \left(1 - \sum_i \alpha_i^l\right) q^{\emptyset} = ev. \tag{1.15}$$

Finally, to make all producer types demand certification we must have in particular

$$f \leq \sum_i \alpha_i^l q^{C^i} + \left(1 - \sum_i \alpha_i^l\right) q^{\emptyset} \tag{1.16}$$

i.e. low quality producers ecpected payoff from certification must be non-negative.[26]

[26]More conditions are required in subgame where all producer types demand certification, but the one presented her is the only required for our proof.

From this we can derive an upper bound on certifier profits:

$$\begin{aligned}
\pi^D(f) = f &\overset{(1.16)}{\leq} \sum_i \alpha_i^l q^{C^i} + \left(1 - \sum_i \alpha_i^l\right) q^{\emptyset} \\
&\overset{(1.15)}{=} ev - e\sum_i (\alpha_i^h - \alpha_i^l)(q^{C^i} - q^{\emptyset}) \\
&\overset{(1.13)}{=} ev - ek'(e) = e(v - k'(e)).
\end{aligned}$$

But $e\,(v - k'(e))$ is the profit from implementing effort level e optimally with a full disclosure rule, therefore we have proven $\pi^D(f) \leq \pi^{FD}$.

□

1.A.3 Proofs of Section 1.5

Proof of Proposition 1.2. In any equilibrium in which Assumption 1.1 holds capture may not take place, since otherwise the beliefs of consumers are not consistent with the behavior of the certifier. Hence, condition (1.5) must be satisfied for all b. As mentioned in the text, certifier profits from deviating $\widehat{\Pi}^{FD}(b|f)$ are largest for b approaching v. Taking this limit yields

$$\begin{aligned}
\lim_{b \nearrow v} \widehat{\Pi}^{FD}(b|f) =& \ (1 - e(f))v + e(f) \cdot \left(f + \delta \Pi^{FD}(f)\right) \\
=& \ (1 - e(f))v + \pi^{FD}(f) + \frac{\delta}{1-\delta} e(f) \pi^{FD}(f) \\
=& \ (1 - e(f))v - \frac{\delta}{1-\delta}(1 - e(f))\pi^{FD}(f) + \Pi^{FD}(f).
\end{aligned}$$

Condition (1.5) is thus equivalent to

$$(1 - e(f))v \leq \frac{\delta}{1-\delta}(1 - e(f))\pi^{FD}(f).$$

Rearranging this expression yields that condition (1.5) is satisfied if and only if

$$\delta \geq \delta^{FD}(f) \equiv \frac{v}{v + \pi^{FD}(f)}.$$

□

Proof of Proposition 1.3. We first argue how condition (1.6) can be translated towards (1.8). Recall $\pi^{FD}(f) = e(f) \cdot f$ and optimal investment by producers requires $k'(e) = v - f$. Replacing f by $v - k'(e)$ yields (1.8). All other statements are straightforward reformulations of Proposition 1.2 and Corollary 1.1. $\square$

Proof of Proposition 1.4. In any equilibrium in which Assumption 1.2 holds capture may not take place, since otherwise the beliefs of consumers are not consistent with the behavior of the certifier. Hence, condition (1.9) must be satisfied for all b. We compute the respective critical discount factors. Taking the limit of $\widehat{\Pi}^D(b|f)$ as b approaches $f + (1-\alpha)(v - V_2)$ we get

$$\begin{aligned}\lim_{b \nearrow f+(1-\alpha)(v-V_2)} \widehat{\Pi}^D(b|f) &= f + (1-\alpha)(v - V_2) + e(\alpha)\delta\Pi^{PD}(f) \\ &= f + (1-\alpha)(v - V_2) + e(\alpha)\frac{\delta}{1-\delta}f \\ &= (1-\alpha)(v - V_2) - \frac{\delta}{1-\delta}(1 - e(\alpha))f + \Pi^{PD}(f).\end{aligned}$$

Consequently this limit lies below $\Pi^{PD}(f)$ if and only if

$$(1-\alpha)(v - V_2) \leq \frac{\delta}{1-\delta}(1 - e(\alpha))f,$$

respectively whenever

$$\delta \geq \delta^{l,h}(\alpha, f) = \frac{(1-\alpha)(v - V_2)}{(1-\alpha)(v - V_2) + (1 - e(\alpha))f}.$$

Similarly the limit of $\widehat{\Pi}^D(b|f)$ as b approaches $f + (v - V_2)$ can be rewritten as follows

$$\begin{aligned}\lim_{b \nearrow f+(v-V_2)} \widehat{\Pi}^D(b|f) &= (1 - e(\alpha)) \cdot (f + (v - V_2)) + e(\alpha)\big(f + \delta\Pi^{PD}(f)\big) \\ &= (1 - e(\alpha))(v - V_2) - \frac{\delta}{1-\delta}(1 - e(\alpha))f + \Pi^{PD}(f).\end{aligned}$$

Therefore $\lim_{b \nearrow f+(v-V_2)} \widehat{\Pi}^D(b|f) \leq \Pi^{PD}(f)$ if and only if

$$\delta \geq \delta^l(\alpha, f) = \frac{v - V_2}{f + v - V_2}.$$

Because capture-proofness requires $\widehat{\Pi}^D(b|f) \leq \Pi^{PD}(f)$ for all b, (1.10) follows. $\square$

Proof of Corollary 1.2. As discussed in the text, the certifier may set $f = V_2$ to minimize the threat of capture. We consider $\delta^l(\alpha, f)$ first. Making use of $f = V_2$ allows us to simplify it to $(v - V_2)/v$. From the proof of Proposition 1.1 we get $V_2 = e(v - k'(e))$ and therefore

$$\delta^l(\alpha) = \frac{v - V_2}{v} = \frac{v - e(\alpha)(v - k'(e(\alpha)))}{v}.$$

Now consider $\delta^{l,h}(\alpha, f)$. With $f = V_2$ we may rewrite

$$\delta^{l,h}(\alpha, f) = \frac{(1-\alpha)(v - V_2)}{(1-\alpha)(v - V_2) + (1 - e(\alpha))V_2}$$

By Bayesian updating we have $V_2 = v \cdot \big((1-\alpha)e(\alpha)\big)/\big(1 - \alpha e(\alpha)\big)$ in equilibrium, which implies $v - V_2 = v \cdot \big(1 - e(\alpha)\big)(\big(1 - \alpha e(\alpha)\big)$. Replacing V_2 and $v - V_2$ accordingly yields

$$\frac{(1-\alpha)(v - V_2)}{(1-\alpha)(v - V_2) + (1 - e(\alpha))V_2} = \frac{1}{1 + e(\alpha)}.$$

$\square$

Proof of Proposition 1.6. Recall, that with full disclosure the critical discount factor is $\delta^{FD}(e) = \frac{v}{v + \pi^{FD}(e)} = \frac{v}{v + e(v - k'(e))}$ and this term is minimized for the profit maximizing effort e, yielding $\min_e \delta^{FD}(e) = \frac{v}{v + \pi^{FD}}$. For all $e \in (0, e^*)$ we have $\frac{v - e(v - k'(e))}{v} < \delta^{FD}(e)$. To see this:

$$\frac{v - e(v - k'(e))}{v} < \delta^{FD}(e) = \frac{v}{v + e(v - k'(e))} \quad \Leftrightarrow \quad \left(e(v - k'(e))\right)^2 > 0.$$

Also

$$\frac{1}{1+e} < \delta^{FD}(e) = \frac{v}{v + e(v - k'(e))} \quad \Leftrightarrow \quad ek'(e) > 0$$

Therefore also $\max\{\frac{1}{1+e}, \frac{v - e(v - k'(e))}{v}\} < \delta^{FD}(e)$ for all $e \in (0, e^*)$ and hence we can define

$$\delta^{PD} := \min_e \max\left\{\frac{1}{1+e}, \frac{v - e(v - k'(e))}{v}\right\}$$

and it follows that $\delta^{FD} > \delta^{PD}$.

Since both $\delta^{PD,l}(e) < \delta^{FD}(e)$ and $\delta^{PD,l,h}(e) < \delta^{FD}(e)$ the last statement follows immediately. $\square$

Proof of Proposition 1.7. When $\delta^{PD} < \tilde{\delta}(\pi^{FD})$ we must have

$$\tilde{\delta}(\pi^{FD}) = \delta^{PD}(e^{FD}) = \frac{1}{1 + e^{FD}}.$$

Because $1/1 + e$ decreases in e we have $\delta^{PD}(e) > \tilde{\delta}(\pi^{FD})$ for any $e < e^{FD}$. Consequently, we must have $\delta^{PD}(e) < \tilde{\delta}(\pi^{FD})$ on some interval $[e^{FD}, \hat{e}]$. This proves our result. $\square$

1.B Extensions

1.B.1 An example with more than two levels of quality

Example 1.1.

Let quality levels be $\{0, 0.5, 1\}$ and $P(q = 0.5|e) = P(q = 1|e) = e/2$. Consequently $P(q = 0|e) = 1 - e$. The cost of effort is $k(e) = e^2/2$. If we restrict the analysis to deterministic disclosure rules, it is straightforward to show that full disclosure with a fee $f = 3/8$ maximizes certifier profits. With this fee both quality levels 0.5 and 1 get certified in equilibrium. Using the same line of argument as in the main text, this disclosure rule can be sustained as a capture-proof equilibrium whenever $\delta \geq \frac{16}{19}$.

A cut-off disclosure rule that certifies any product whose quality exceeds 0, but does not distinguish any further, achieves the same static profit as the mentioned full disclosure rule. However, the largest possible bribe is then not equal to 1 since no certificate which yields a price of one is available. Instead, the best certificate yields $3/4$, the value of a certified product. Consequently, a capture-proof equilibrium with this disclosure rule exists whenever $\delta \geq \frac{16}{20}$. While profits remain the same, the largest acceptable bribe is lowered.

1.B.2 Alternative disclosure rules for the two-quality case

Proposition 1.8. *For any $\delta < \delta^{FD}$ and any disclosure rule which is such that the highest certificate's value is different from v, no capture-proof equilibrium exists.*

Proof. We restrict the proof to the following simple disclosure rule[27]: there are two certificates, C_1 and C_2, where high quality always receives C_1 and low quality receives

[27] For any other rule, the argument is the same for selling the best certificate in a capture offer to the

C_1 with probability $\alpha \in (0,1)$. Denote V the value of C_1, certificate C_2 is always worth zero (in equilibrium). The first-order condition for producer investment reads as

$$k'(e) = (1-\alpha)V$$

and from Baye's rule we have

$$V = v\frac{e}{e+\alpha(1-e)}.$$

Thus, to implement a particular e, the certifier has to set[28]

$$\alpha = \frac{e(v-k'(e))}{e(v-k'(e))+k'(e)}$$

The fee must be such that low quality producers are willing to get their product certified, i.e. $f \leq \alpha V$.

When a purchased product with certificate C_1 turns out to be of low quality, consumers cannot be sure whether this was due to bad luck or to a captures certifier. Appropriate trigger beliefs have to be such that the certifier is punished whenever low quality is sold with certificate C_1. This can well happen without any deviation by the certifier. The probability of entering punishment, absent any deviation, is $p = (1-e)\alpha$ and expected profits from honest play are given by

$$\Pi^h(\alpha, f) = f + (1-p)\delta f + (1-p)^2\delta^2 f + \ldots = \frac{f}{1-(1-p)\delta}.$$

The maximal bribe is given by $b \approx (1-\alpha)V + f$, where only low quality producers accept it. The profit from making such an offer is

$$\Pi(b|f,\alpha) = (1-e)b + e(f + \delta\Pi^h(\alpha, f))$$

We have $\Pi(b|f,\alpha) \leq \Pi^h(\alpha, f)$ for $b \to (1-\alpha)V + f$ whenever

$$\delta \geq \frac{(1-e)b-(1-e)f}{(1-e)(1-p)b-epf} = \frac{b-f}{\big(1-(1-e)\alpha\big)b - e\alpha f}$$

low quality producer. However, there are even more feasible bribing offers, which make it even harder to resist the threat of capture.

[28]Note that $\lim_{e\to 0}\alpha$ equals 1 whenever $k''(0) = 0$ and otherwise equals $\frac{v}{v+k''(0)\in(0,1)}$, that is in the latter case not all α are implementable.

This is both increasing in b and in f, such that the largest threat is exercised for $f = \alpha V$ and $b = V$, which results in the condition

$$\delta \geq \frac{1}{1 + e\alpha}.$$

We have $\frac{1}{1+e\alpha} \geq \frac{v}{v+e(v-k'(e))}$ if and only if

$$v - k'(e) \geq v\alpha \quad \Leftrightarrow \quad 1 \geq e.$$

Hence, for all e to be implemented, this is only possible with a noisy rule without sure high quality certificate, when this is also possible using a full disclosure rule.

□

Chapter 2

Mediated Audits

2.1 Introduction

The beneficial role of audits in mitigating incentive problems has been widely acknowledged.[1] Findings of an audit are used to punish misbehavior, which dampens incentives for misreporting and permits a reduction of informational rents. In this respect, commitment plays a crucial role, because the threat of a punishment is only effective if it is credible. Under the assumption of full commitment, however, agents report truthfully and an audit never leads to a penalty payment. Conducting a costly audit is therefore not in the principal's interest ex-post. As a consequence, commitment to an audit policy, i.e., a contractually fixed probability of an audit conditional on observable and verifiable information, seems implausible.[2]

In this chapter I study the optimal contract when costly audits are available, but it is impossible to commit to an audit strategy.[3] Usually, the revelation principle applied to principal-agent problems with full commitment implies that the principal offers a menu of contracts containing one item per type of agent. The principal's inability to credibly commit to an audit strategy leads to the undermining of this simple structure. Rather, the principal resorts to an impartial mediator. The mediator breaks the infor-

[1]See for example, Baron (1984), Baron (1989), Baron and Besanko (1984), Border and Sobel (1987), Demski, Sappington, and Spiller (1987), Dunne and Loewenstein (1995), Graetz, Reinganum, and Wilde (1986), Hart (1995), Kofman and Lawarrée (1993) and Mookherjee and Png (1989).

[2]Baron and Besanko (1984), Border and Sobel (1987) and Kofman and Lawarrée (1993), among others, assume commitment to an audit strategy.

[3]Other authors have studied auditing with limited commitment, e.g., Khalil and Lawarrée (1995), Khalil (1997), Khalil and Parigi (1998), Khalil and Lawarrée (2006), Graetz, Reinganum, and Wilde (1986), Melumad and Mookherjee (1989) and Chatterjee, Morton, and Mukherji (2008). All these articles make restrictive assumptions on the communication, which are relaxed in this paper.

mation flow between agent and principal, and, in particular, conveys only as much information about the agent's type as is required to persuade the principal to audit. For a large set of parameters the optimal contract entails strictly more contingencies than types, and can not be implemented with simple communication protocols.

The structure of the optimal mechanism improves our understanding of the institutional design of audit agencies. A separation of auditing and contracting is inherent in the optimal mechanism and should therefore also be reflected in the institutions. Separate audit agencies are just one way of implementing this idea. The analysis of this chapter also provides a proper benchmark to which other attempts in mitigating the commitment problem can be compared to.[4] Furthermore, I completely characterize the optimal communication mechanism in a setting with limited commitment without restrictions on the communication. The analysis reveals new insights on the beneficial role of mediation, and advances the quest for a unified and tractable model to study problems of contracting with limited commitment.

As a workhorse, I use a two type-version of the well-known model of regulating a monopolist with unknown cost, introduced by Baron and Myerson (1982) and add the possibility of an ex-post audit. To account for the lack of commitment, the principal has to resort to general coordination mechanisms, in the spirit of Myerson (1986). The crucial difference to the full commitment case is that communication between principal and agent is in general neither one-shot nor face-to-face. Instead, the parties resort to an impartial mediator, a machine or some noisy channel.

With limited commitment the contracting game essentially consists of two stages: First a negotiation stage, where production schedules are agreed upon, and second the audit stage. For an audit to take place, the posterior the principal holds in stage two must be such that he expects to earn more from conducting an audit than the costs of an audit.

The optimal mediated contract has the following properties: (1) the agent reports truthfully to the mediator, (2) the mediator performs a report-dependent randomization, (3) the randomization is over a report-dependent transfer-quantity schedule and a fixed transfer-quantity schedule, (4) only the fixed transfer-quantity schedule is accompanied with a recommendation to audit, and (5) the principal obediently follows the mediator's recommendation. Properties (1) and (5) follow from applying the revelation principle. (2) is used to guarantee the principal's obedience after a recommen-

[4]For instance Khalil and Lawarrée (2006) contrast internal versus external auditors, but use an unmediated mechanism for internal audits as the benchmark.

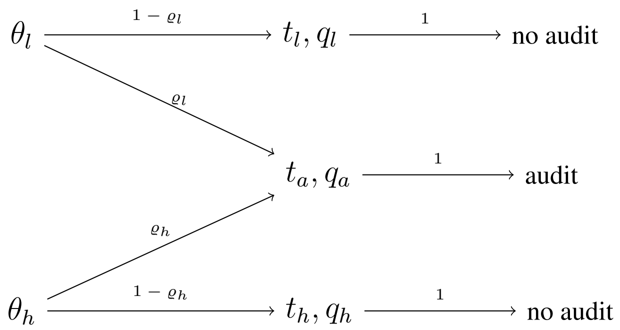

Figure 2.1: The optimal coordination mechanism with limited commitment.

dation to audit. In particular, it creates the right posterior that makes the principal indifferent between auditing and not when the recommended action is audit. The exact details of the three utilized transfer-quantity schedules is determined by trading off rents, efficiency and audit costs.

The structure of the optimal mechanism is illustrated in Figure 2.1. The introduction of the transfer-quantity schedule (t_a, q_a) distinguishes it from a traditional self-selection menu. Audits are only recommended in combination with (t_a, q_a). To this end, the respective probabilities assigned to this outcome are linked via the principal's obedience constraint. The share of low-cost types is such, that the principal is indifferent between following the recommendation and not.

A crucial feature of the mediated mechanism is the confidential randomization. When receiving a recommendation to audit, the principal does not know which type triggered this recommendation. This distinguishes the mediated mechanism from a random contract, where the randomization is performed by the principal. In the latter, the principal learns the agent's reported information and uses it for any subsequent action. In the proposed mediated mechanism the mediator fine-tunes the information flow between agent and principal. In particular, three different posteriors are generated: two degenerated posteriors and one interior which triggers the audit.

Whether the optimal outcome features all three transfer-quantity schedules depends on the level of the deterrent. For low levels of the deterrent audits are not profitable and, hence, never recommended. When the deterrent is moderate, audits are used to reduce the rent that is left to the low-cost type. As long as it remains impossible to reduce this rent to zero, audit frequency is at its maximal level. This in particular implies that all high-cost types are audited, i.e. are assigned (t_a, q_a) with probability one. The schedule (t_h, q_h) is, consequently, not produced *in* equilibrium, but plays a crucial role in determining whether audits are profitable in the first place. The classical rent vs. efficiency trade-off is still in place, but it is softened because the efficient type produces under the schedule (t_a, q_a) with strictly positive probability.

When the rent can be easily reduced to zero, the principal starts reducing the ex-ante probability of an audit and therefore the induced audit-cost. Each type now produces the intended pair (t_r, q_r) more frequently. The classical rent-vs-efficiency trade-off is augmented to a rent-vs-efficiency-vs-audit-cost trade-off. Increasing q_h, resp. q_a, has on the one hand the classical rent-effect: it increases the rent that is left to the low-cost type. On the other hand, audits are used to reduce this rent to zero and therefore more audits are required. In particular, the allocation has to be distorted more via a larger likelihood of q_a - the audit-cost-effect. The interaction of these two effects for each individual quantity and across the two quantities determines the optimal communication mechanism.

Comparing the results to the full commitment audit contract, it can be shown that audits get profitable only for larger deterrents and it requires a larger deterrent to reduce the agent's rent to zero. However, some comparative statics are similar. In particular, as the deterrent increases the outcome converges to the first-best and the ex-ante probability of an audit converges to zero.

The structure of the optimal mechanism has implications into two directions: From a practical perspective it provides a novel explanation for the frequently observed separation of contracting and auditing.[5] With limited commitment this separation is an inherent feature of the optimal contract and separate agencies are just one way of implementing the optimum. In particular the optimal mechanism requires to break down the flow of information into two parts: one that is needed to assign the production schedule and one that is needed to decide upon audits.

From a theoretical perspective, the chapter provides a complete characterization of the optimal contract under limited commitment. The optimal contract exhibits interesting features that help understanding the beneficial role of mediation in contract theory with limited commitment. The mediated mechanism correlates the agent's reported information with the recommendation to the principal. The aspect of correlation can already be found in Myerson (1982). However, it plays no role in principal-agent problems with only one agent and full commitment. This is maybe a reason

[5]Examples for this separation abound: U.S. Government Accountability Office (GAO), Internal Affairs Bureau in the police force and Internal Audit Service in the European Commission.

why it has not yet been fully acknowledged in the literature on contracting with limited commitment.

The remainder of this chapter is organized as follows: Section 2.2 reviews the related literature and section 2.3 presents the model. Section 2.4 reviews the optimal contract under full commitment. In section 2.5 the optimal coordination mechanism under limited commitment is analyzed. Section 2.6 concludes. All proofs are relegated to the appendix.

2.2 Related literature

This article contributes to the literature on auditing with limited commitment. Khalil (1997) studies a model similar to the one studied here, but limits the analysis to one-shot face-to-face communication. The mixed equilibria obtained by Khalil do never yield optimal contracts. Graetz, Reinganum, and Wilde (1986) have modeled tax compliance without commitment. However, their model is not one of mechanism design, but a game-theoretic analysis of an inspection game. Dunne and Loewenstein (1995) study a model where agents compete for a principal's project. The principal can observe the agent's cost at some private expense, but cannot credibly commit to actually do so. Chatterjee, Morton, and Mukherji (2008) study a continuous-type model, but focus on one-shot face-to-face communication.

Also the literature on costly state verification studies the issue of limited commitment, nicely surveyed by Attar and Campioni (2003). In these models contracting typically takes place *before* the agent receives private information, e.g., Khalil and Parigi (1998) study loan contracts where the bank can audit in case of default. Or contracts arise in a competitive market, e.g., Picard (1996) studies optimal insurance contracts when insurers cannot commit to audit strategies.

Limited commitment has been used to explain the frequently observed separation between contracting and auditing. Khalil and Lawarrée (2006) derive a demand for external audits when internal commitment is limited. Melumad and Mookherjee (1989) study delegation in a model of tax compliance. The government can implement the full-commitment solution by delegating authority over the audit policy.

Surprisingly, little is known about optimal contracts under limited commitment in general. Bester and Strausz (2001) provide a revelation principle when communication is limited to one round of face-to-face communication. The same authors show that noisy communication can be beneficial, in Bester and Strausz (2007). Their results depend on an assumption about the agent's utility function, which is not satisfied in this chapter. Therefore, we cannot adopt their solution procedure.

The general approach to communication used in the chapter is borrowed from the game-theoretic literature, e.g., Myerson (1986) and Forges (1986). That multistage communication already enhances welfare has been demonstrated by, e.g., Forges (1990) and Krishna and Morgan (2004). With indirect communication, i.e., via a mediator or a noisy channel, further improvements are possible (see, e.g., Myerson (1986) and Forges (1986)).

Recently, mediation has found its way into contract theory. Rahman and Obara (2010) show that mediation can virtually implement first-best effort choices in a team problem where budget-balance is required. Strausz (2012) links this result to general insights from mechanism design.

The impact of various communication protocols is perhaps best understood in the area of cheap talk. Crawford and Sobel (1982) provide the benchmark with a single round of face-to-face communication, which is extended to multiple rounds by Krishna and Morgan (2004). Mediation is added by Goltsman, Hörner, Pavlov, and Squintani (2009). That noisy communication is already sufficient to achieve the outcome under mediation, is shown in Blume, Board, and Kawamura (2007). Also Mitusch and Strausz (2005) study beneficial mediation in a cheap talk model of conflict.

2.3 Model

Consider the following principal-agent framework: A principal hires an agent to carry out the production of some good. The principal's value of q units of this good is given by the strictly increasing and strictly concave function $V(q)$. We normalize $V(0) = 0$ and to guarantee strictly positive, but bounded output levels, we shall assume $V'(0) = \infty$ and $V'(\infty) = 0$. The agent has constant marginal costs of production, given by the parameter $\theta > 0$. At the outset, only the agent knows his marginal cost. With probability $\phi \in (0, 1)$ marginal costs are low, i.e., $\theta = \theta_l$, and with probability $1 - \phi$ marginal costs are high, i.e., $\theta = \theta_h > \theta_l$. In particular, $\Delta\theta := \theta_h - \theta_l > 0$. Efficient production levels are thus given by

$$V'(q_i^o) = \theta_i, \quad i = l, h. \tag{2.1}$$

For delivery of a publicly observable quantity q, the principal pays the agent a monetary transfer t. In addition, we assume the principal possesses an audit technology, that allows to learn the agent's true marginal cost after production took place.[6] As an example, the principal can send an inspector in order to check the agent's accounts after production. From the accounts the inspector can infer total production costs and thereby learn the true parameter θ.

An audit is costly to the principal and this cost is given by $c > 0$. The result of the audit can be used to demand a payment from the agent to the principal, which is assumed to be exogenously fixed at a level $P > 0$.[7] Throughout we use the notation *penalty scheme* for a mapping $\mathcal{P} : \{\theta_l, \theta_h\} \to \{0, P\}$, which gives the true cost types that are required to make an adjustment payment under the current regime. In total, there are four possible penalty schemes: the void penalty scheme $\mathcal{P}^0 \equiv 0$, the complete penalty scheme $\mathcal{P}^{lh} \equiv P$ and then one for each type, i.e. $\mathcal{P}^i$ satisfies $\mathcal{P}^i(\theta_i) = P$ and $\mathcal{P}^i(\theta_j) = 0$ for $j \neq i$. Lastly, denote α the audit strategy of the principal, i.e. $\alpha \in [0, 1]$ is the probability with which the principal conducts an audit. The payoff of an agent of type θ_i, when producing quantity q, with transfer t and penalty scheme $\mathcal{P}(\cdot)$, when the principal's audit strategy is α, is given by[8]

$$U_i = t - \theta_i q - \alpha \mathcal{P}(\theta_i). \tag{2.2}$$

The principal's payoff in that case is

$$V_i = V(q) - t + \alpha\big[\mathcal{P}(\theta_i) - c\big]. \tag{2.3}$$

This essentially describes a two-stage game, where in stage one the agent produces quantity q and transfers are paid, and in stage two the principal can audit and adjustment payments are made. It is natural to assume the principal wants to use information

[6] The assumption of a perfect audit technology simplifies the analysis, but is not crucial for our results.

[7] One possible interpretation is, that P is enforced by a court. Though the contracting parties have all flexibility in determining conditions that are seen as breach, they are committed to penalty payments imposed by jurisdiction.

[8] From this formulation of the agent's utility it can be easily seen that Assumption 1 in Bester and Strausz (2007) is not satisfied. With the interaction of production costs and penalty scheme a single-crossing property is precluded. Effectively the problem is one of multi-dimensional screening, with a cost-type and an audit-type, though substantially they are identical. What complicates the problem is that screening the second-dimension is endogenous, i.e. depends on the penalty scheme that itself is part of the contract.

from the first stage when deciding on her strategy in the second stage. We thus assume that the principal cannot commit to her *audit strategy*.

More formally, in the first stage the principal can commit to a transfer t, a quantity q and penalty scheme $\mathcal{P}(\cdot)$. In the second stage she chooses the audit strategy α. The audit strategy is not contractible in stage one, so the principal chooses it at her own discretion. Before formally introducing coordination mechanisms, we briefly analyze the relevant benchmark cases that have been extensively studied in the literature.

2.4 Benchmarks

This section considers two important benchmark cases: First we briefly review results on optimal contracts in our setting, when no audit technology is available. Second, we analyze the case of full commitment, where the principal can commit to an audit strategy.

2.4.1 No audits

Our setting without audits is a special case of the well-known framework of regulating a monopolist with unknown costs, established by Baron and Myerson (1982). Applying the revelation principle, the following structure is without loss: The agent is asked to report his cost type to the principal who *commits* to terms of trade for each type report. Let (t_i, q_i) be the contract executed when the agent reports to be of type θ_i. Baron and Myerson (1982) have shown that the following is optimal.

Lemma 2.1. *The optimal contract without audits is given by*

$$q_l^{na} = q_l^o, \qquad V'(q_h^{na}) = \theta_h + \frac{\phi}{1-\phi}\Delta\theta. \tag{2.4}$$

In equilibrium, $U_h^{na} = 0$ *and* $U_l^{na} = \Delta\theta q_h^{na}$.

The optimal contract trades off the rent that has to be left to the efficient type and the productive distortion imposed on the inefficient type: A higher quantity q_h increases the rent left to the efficient type, but increases profits from the inefficient type's production.

2.4.2 Full commitment

We now introduce the audit technology, but make the assumption that the principal can also commit to an audit strategy. The sequential structure, i.e., that audits are conducted *after* production is therefore inconsequential. Applying once more the revelation principle, it is without loss that the principal offers a menu

$$\Gamma = \{(t_l, q_l, \mathcal{P}_l, \alpha_l), (t_h, q_h, \mathcal{P}_h, \alpha_h)\}, \tag{2.5}$$

consisting of a transfer, a quantity, a penalty scheme and an audit strategy for each potential type report. The optimal contract solves the following problem:

$$\max_{\Gamma} \ \phi\big(V(q_l) - t_l + \alpha_l(\mathcal{P}_l(\theta_l) - c)\big) + (1-\phi)\big(V(q_h) - t_h + \alpha_h(\mathcal{P}_h(\theta_h) - c)\big) \tag{2.6}$$

subject to the following constraints

$$\begin{aligned} t_l - \theta_l q_l - \alpha_l \mathcal{P}_l(\theta_l) &\geq 0 & (2.7)\\ t_h - \theta_h q_h - \alpha_h \mathcal{P}_h(\theta_h) &\geq 0 & (2.8)\\ t_l - \theta_l q_l - \alpha_l \mathcal{P}_l(\theta_l) &\geq t_h - \theta_l q_h - \alpha_h \mathcal{P}_h(\theta_l) & (2.9)\\ t_h - \theta_h q_h - \alpha_h \mathcal{P}_h(\theta_h) &\geq t_l - \theta_h q_l - \alpha_l \mathcal{P}_l(\theta_h) & (2.10) \end{aligned}$$

where (2.7) and (2.8) are the respective type's participation constraints, and (2.9) and (2.10) are the incentive constraints. The solution to this problem has been studied in Baron and Besanko (1984) and Kofman and Lawarrée (1993).[9]

Lemma 2.2. *The optimal contract under full commitment coincides with the no-audit contract (see Lemma 2.1) if and only if* $\phi P \leq (1-\phi)c$.[10] *Otherwise, the audit contract entails* $\alpha_l^c = 0, q_l^c = q_l^o, U_h^c = 0$ *and the penalty schemes are* $\mathcal{P}_h^c = \mathcal{P}^l$ *and* $\mathcal{P}_l^c = \mathcal{P}^0$. *If* $P \leq \underline{P}^c := \Delta\theta q_h^{na}$ *we further have* $\alpha_h^c = 1, q_h^c = q_h^{na}$ *and* $U_l^c = \Delta\theta q_h^{na} - P$. *For* $P > \underline{P}^c$ *the audit contract entails* $U_l^c = 0$ *and either*

- $\alpha_h^c = 1$ *and* $q_h^c = P/\Delta\theta$*, or*
- $\alpha_h^c < 1$ *and* $V'(q_h^c) = \theta_h + \frac{c}{P}\Delta\theta$*, as well as* $\alpha_h^c = \Delta\theta q_h^c/P$.

[9] Both papers use similar, though not identical, models. We adopt a different solution procedure, that allows for finer results, see Corollary 2.1.

[10] Here and in the following we adopt the following convention: Audits are only used when the principal *strictly* prefers to do so. Otherwise there is an indeterminacy for the unique level of P where the principal obtains the same profit from any level of auditing.

This illustrates the impact of audits on the optimal contract. Without auditing, two inefficiencies arise: a rent has to be paid to the efficient type and the inefficient type's production is distorted. The principal uses the audit technology to gradually reduce these distortions. First of all, it is immediate that via the penalty P the rent to the efficient type can be reduced by P. The gain is ϕP, but the cost is $(1-\phi)c$ - the principal now has to audit the inefficient type to establish the deterrent. Clearly, this is beneficial if and only if $\phi P > (1-\phi)c$.[11] If, given the allocation, the penalty is large enough to reduce the efficient type's rent to zero, the principal also reduces the allocative distortion. Yet, for any finite penalty P there remains a distortion on q_h - only in the limit as P approaches infinity does the allocation converge to the first-best.

Whether $\alpha_h^c < 1$ for $P > \underline{P}^c$ depends on the shape of V. Corollary 2.1 gives a general result under a mild assumption on the function V.

Corollary 2.1. *Assume* $-\frac{V''(q)q}{V'(q)} \geq (\phi\Delta\theta)/((1-\phi)\theta_h + \phi\Delta\theta)$. *Then there exists a unique value* $\overline{P}^c < \Delta\theta q_h^o$ *such that* $\alpha_h^c < 1$ *if and only if* $P > \overline{P}^c$. *Moreover,* α_h^c *strictly decreases for* $P > \overline{P}^c$ *and* q_h^c *strictly increases for* $P > \underline{P}^c$.

Figure 2.1 illustrates the comparative statics of the audit contract with respect to the deterrent P. For low values of P, the principal never audits and the allocation is that from the no-audit contract in Lemma 2.1. Next, there is a range of values P, where (sure) audits are used to reduce rents. As soon as P is large enough to reduce all rents to zero, distortions on quantities can be reduced. On a first range this is done with certain audits. As P gets sufficiently large, i.e., for $P > \overline{P}^c$, audits are random and their frequency decreases with P, while the allocation converges to first-best. However, for any finite P the allocation is distorted.

2.5 Limited commitment

This section studies optimal contracts when commitment to an audit strategy is impossible. To begin with, it is instructive to point out that the outcome from the contract of Lemma 2.2 cannot be implemented when commitment to an audit strategy is impossible. To see this, notice that because the agent reports truthfully the principal knows that the agent is of type θ_i, whenever the reported type is θ_i. Consequently, conducting an audit after report θ_h never results in the collection of the penalty P, and

[11] Notice that this does not even require $P > c$, i.e. for large ϕ a relatively small penalty compared to a high cost of audit is sufficient for its profitability.

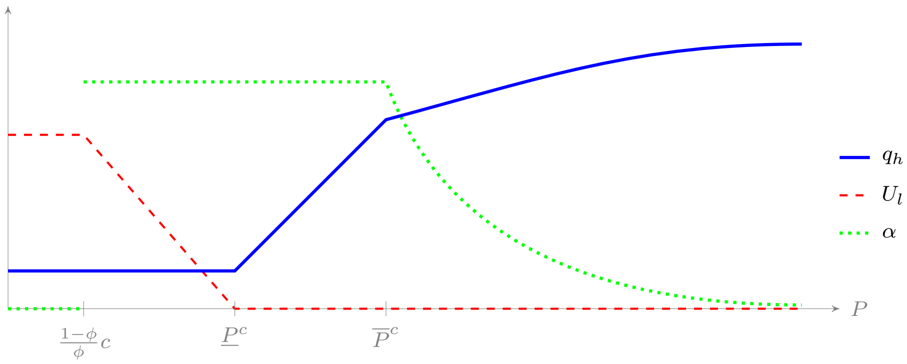

Figure 2.1: Comparative statics of the audit contract under full commitment as P changes.

hence the principal has no incentive to audit at all. But if the principal does not audit, incentive compatibility fails: the θ_l-type now strictly prefers to report θ_h.

When studying the optimal contract under limited commitment, we have to take the sequential structure into account: In the first stage principal and agent bargain over the contractual variables t, q and the penalty scheme. In the second stage the principal decides upon the audit, taking into account all information revealed in stage one.

Applying the revelation principle for multistage games with communication, see Myerson (1986), the principal can confine attention to incentive compatible communication mechanisms. A (direct) communication mechanism is defined as follows: First, the agent reports all his private information to some central mediator; then the mediator computes an allocation $(q, t, \mathcal{P})$ and a recommended action for the principal, as a (potentially random) function of the report; the allocation gets implemented, i.e., the agent produces q and receives the transfer t, and the mediator confidentially tells the principal what is her recommended action. Such a communication mechanism is incentive compatible, if the agent finds it in his best interest to truthfully report his information, and if the principal cannot gain by disobeying the mediator's recommendation.

Let $\boldsymbol{\pi} = (\pi_l, \pi_h)$ be a vector of probability distributions, where $\pi_i(t, q, \mathcal{P}, r)$ denotes the probability that outcome $(t, q, \mathcal{P})$ is chosen and the principal's recommended action is r, after the agent reported to be of type θ_i. The difficulty in finding the optimal incentive compatible communication mechanism lies in finding the support of the probability distributions π_i.[12] In order to avoid measure-theoretic complications we

[12]In the static counterpart this issue is circumvented by focussing on *deterministic* mechanisms, hence the respective support is a singleton. The crux of the matter is that an optimal contract that

shall assume countability of the support of both π_l and π_h.

To simplify the exposition, we formally set up the principal's problem of finding the optimal communication mechanism only in Appendix 2.A.2, where we prove the following statement that substantially simplifies the following analysis. Let $r = a$ be the recommendation to audit, and $r = na$ the recommendation not to audit.

Proposition 2.1. *Without loss of generality an optimal communication mechanism has the following properties: Each π_i randomizes over $(t_i, q_i, \mathcal{P}^0, na)$ and $(t_a, q_a, \mathcal{P}^l, a)$. The incentive constraint of type θ_l as well as the participation constraint of type θ_h are binding, while the incentive constraint of type θ_h is slack. Moreover, the principal is indifferent between auditing and not after receiving the recommendation to audit.*

This result is essentially obtained as follows: First we adopt the standard procedure of 'guessing' that the high-cost type's incentive constraint is not binding. This allows us to rule out penalties for type θ_h, as they require larger transfers to this type. Via the efficient type's incentive constraint also the transfer to type θ_l is then increased. Second, obedience only matters when the recommendation is to audit, because otherwise setting the penalty scheme to $\mathcal{P}^0$ makes this trivial. Quantities that are produced by type θ_i without the recommendation to audit must be unique, because otherwise it is not worse to put all mass on one of them that yields the highest surplus.[13] Lastly, the obedience constraint must be binding, which also yields uniqueness of the associated quantity.

Figure 2.1 illustrates the optimal communication mechanism, using the results from Proposition 2.1. To simplify notation, let ϱ denote the probability that (t_a, q_a) is chosen by the mediator after the agent reported to be of type θ_h, i.e.

$$\varrho := \pi_h(t_a, q_a, \mathcal{P}^l, a). \tag{2.11}$$

To keep the principal indifferent whether to audit, it has to hold that $\phi^a P = c$, where ϕ^a denotes the principal's posterior after this recommendation. Solving this equation yields[14]

$$\pi_l(t_a, q_a, \mathcal{P}^l, a) = \varrho \frac{(1-\phi)c}{\phi(P-c)}. \tag{2.12}$$

prescribes audits is never deterministic.

[13]This argument uses strict concavity of V. It is similar to arguing that the optimal contract under full commitment is deterministic.

[14]At this point we might still have $\pi_l(q_a, P_l, a) > 1$. We show in Proposition 2.2 below, that $\phi P > c$ whenever $\varrho > 0$. This implies that all probabilities are indeed between zero and one.

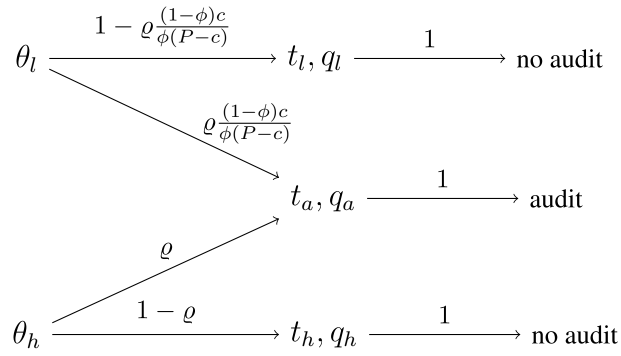

Figure 2.1: The optimal coordination mechanism with limited commitment.

Feasibility further requires $\pi_h(t_h, q_h, \mathcal{P}^0, na) = 1 - \varrho$, and $\pi_l(t_l, q_l, \mathcal{P}^0, na) = 1 - \varrho\frac{(1-\phi)c}{\phi(P-c)}$.

The variable ϱ represents the audit probability in the communication mechanism. Obviously, whenever $\varrho = 0$ no audits are conducted. In what follows, $\varrho = 1$ represents the maximal audit probability - all high-cost types are audited and some low-cost types. In the latter two cases only two transfer-quantity pairs occur in equilibrium. The third case refers to $\varrho < 1$, when all three transfer-quantity pairs occur. For this case it is essential that the principal does not observe the agent's report and holds a non-trivial belief after being recommended to audit. Due to limited commitment, the mechanism has to recommend an audit for some low- and some high-cost types. If the share of high-cost types is too high, the principal will no longer follow the recommendation and obedience fails. If, however, the share of low-cost types is too high it becomes too costly to incentivize the agent to report truthfully. Following Proposition 2.1 it is optimal to keep the principal just indifferent whether to follow the recommendation. With the resulting indifference condition all probabilities but one are fixed, as described above.

The principal's profit from employing the communication mechanism from figure 2.1 is

$$\begin{aligned} &\phi\left(1-\varrho\frac{(1-\phi)c}{\phi(P-c)}\right)\big(V(q_l)-t_l\big)+(1-\phi)(1-\varrho)\big(V(q_h)-t_h\big) \\ &\quad+\left(\phi\varrho\frac{(1-\phi)c}{\phi(P-c)}+(1-\phi)\varrho\right)\big(V(q_a)-t_a\big). \end{aligned} \tag{2.13}$$

Notice that revenues from auditing disappeared: Because the obedience constraint is binding, the principal is indifferent between auditing and not whenever recommended to do so. Hence, audit revenues equal zero. Finding the optimal communication mechanism requires maximization of (2.13) over the variables t_l, q_l; t_a, q_a; t_h, q_h and ϱ. Following Proposition 2.1 the principal has to respect the participation constraint of type θ_h

$$\varrho(t_a - \theta_h q_a) + (1-\varrho)(t_h - \theta_h q_h) = 0, \tag{2.14}$$

the incentive constraint of type θ_l

$$\begin{aligned} \left(1 - \varrho\frac{(1-\phi)c}{\phi(P-c)}\right)(t_l - \theta_l q_l) + \varrho\frac{(1-\phi)c}{\phi(P-c)}(t_a - \theta_l q_a - P) = \\ \varrho(t_a - \theta_l q_a - P) + (1-\varrho)(t_h - \theta_l q_h). \end{aligned} \tag{2.15}$$

and the participation constraint of type θ_l

$$\left(1 - \varrho\frac{(1-\phi)c}{\phi(P-c)}\right)(t_l - \theta_l q_l) + \varrho\frac{(1-\phi)c}{\phi(P-c)}(t_a - \theta_l q_a - P) \geq 0. \tag{2.16}$$

Substituting t_l, t_a and t_h from the former two binding constraints, the principal's profit can be written as a function $\mathcal{V}(\varrho, q_l, q_a, q_h)$:[15]

$$\begin{aligned} \mathcal{V}(\varrho, q_l, q_a, q_h) = {} & \phi\big(V(q_l) - \theta_l q_l - \Delta\theta q_h\big) + (1-\phi)\big(V(q_h) - \theta_h q_h\big) \\ & + \varrho\Bigg\{-\frac{(1-\phi)c}{P-c}\big(V(q_l) - \theta_l q_l - V(q_a) + \theta_l q_a\big) \\ & + (1-\phi)\big(V(q_a) - \theta_h q_a - V(q_h) + \theta_h q_h\big) \\ & - \phi\Delta\theta(q_a - q_h) + \frac{\phi P - c}{P-c}P\Bigg\} \end{aligned} \tag{2.17}$$

The principal now maximizes (2.17) subject to (2.16). From (2.17) the impact of audits becomes clear. When $\varrho = 0$, the principal never audits and the expression coincides with the standard no-audit profit. If audits are to be implemented, the mechanism has to create the right incentives for the principal to actually do so. In order to achieve this, a new outcome is created: the quantity q_a along with the recommen-

[15] Interestingly, from two binding constraints we are able to substitute for three variables. The reason is, that only type-specific transfers matter. In fact, there is a continuum of transfer-pairs t_h, t_a that satisfy the inefficient type's participation constraint. For any such pair there exists a unique t_l such that (2.15) is satisfied.

dation to audit. A particular ratio of low-cost and high-cost types is required to meet the obedience constraint. Introducing this new contingency has several effects on the principal's profit: The respective types now produce q_a instead of q_l, resp., q_h. Starting from the no-audit contract, this leads to a loss when interacting with type θ_l and the sign of the effect when interacting with type θ_h is unclear. Second, introducing audits has an effect on the rent to the efficient type. The latter effect is twofold: the deterrent itself affects the rent as long as the frequency of audits differs for the two possible type reports. But also the production itself has an effect, because a false report now leads to a lottery over q_a and q_h.

The next proposition analyzes profitability of audits and gives a necessary and sufficient condition on the size of deterrent P, that guarantees beneficial audits.

Proposition 2.2. *There exists a unique value P^*, defined as the solution of*

$$\max_{q_l,q_a,q_h} \left\{ \frac{(1-\phi)c}{P-c}\big(V(q_l) - \theta_l q_l - V(q_a) + \theta_l q_a\big) + \phi\Delta\theta(q_a - q_h) \right. \\ \left. -(1-\phi)\big(V(q_a) - \theta_h q_a - V(q_h) + \theta_h q_h\big) - \frac{\phi P - c}{P-c}P \right\} = 0 \tag{2.18}$$

such that the optimal coordination mechanism entails $\varrho^* > 0$ if and only $P > P^*$. In particular, we have $\phi P^* > c$.

There are two important insights from Proposition 2.2. A necessary condition for profitable audits is $\phi P^* > c$. Why is this the case? To guarantee obedience, the principal's posterior ϕ^a when she is recommended to audit has to satisfy $\phi^a P = c$. This requires a particular ratio of high- and low-cost types. In particular, when $\phi P \leq c$ the ratio of low-cost types has to be (weakly) higher than it is in the prior distribution - it has to hold that $\phi^a \geq \phi$. Consequently, a low-cost type is more frequently asked to produce q_a and getting audited than a high-cost type. But this makes it harder to satisfy (2.15), because there are fewer audits when the agent misreports, i.e., fewer penalty payments. Hence, the penalty P itself helps to steer incentives if and only if $\phi P > c$. Incidentally, this condition also guarantees that all probabilities lie between zero and one, whenever $\varrho^* > 0$.

Furthermore, recall that the condition for profitable audits under full commitment was $\phi P > (1-\phi)c$. Because $\phi P^* > c$ we trivially have $\phi P^* > (1-\phi)c$ and audits are thus already profitable under full commitment. Consequently, with limited commitment the principal uses her audit technology only for larger levels of the deterrent than under full commitment. In particular, with the latter an audit may well be unprofitable ex-post even if the principal knew the agent's type. To see this, observe that the condition $\phi P > (1-\phi)c$ does not require $P > c$.

Let us now continue with a discussion of the optimal coordination mechanism. When $P \leq P^*$ we have seen that the solution entails $\varrho^* = 0$. It is then straightforward to verify that it coincides with the optimal no-audit contract. This does, however, not imply that there is no commitment problem, because we may well have $\phi P > (1-\phi)c$. In this sense, limited commitment has a large impact here in that it makes audits completely unprofitable, whereas the principal would have used them under full commitment.

For the remainder assume the deterrent is large enough, i.e. $P > P^*$, such that audits are used in an optimal mechanism. The principal's profit in (2.17) can be rewritten as follows

$$\begin{aligned}\mathcal{V}(\varrho, q_l, q_a, q_h) = (1-\varrho)\Big\{\phi\big(V(q_l) - \theta_l q_l - \Delta\theta q_h\big) + (1-\phi)\big(V(q_h) - \theta_h q_h\big)\Big\} \\ + \varrho\Big\{\frac{\phi P - c}{P - c}\big(V(q_l) - \theta_l q_l - \Delta\theta q_a + P\big) + \frac{(1-\phi)P}{P-c}\big(V(q_a) - \theta_h q_a\big)\Big\}\end{aligned} \tag{2.19}$$

Ignoring (2.16) for the moment, the formulation of principal's profit in (2.19) already tells us something about the optimal quantities. Both arise from a classical rent-vs-efficiency trade-off, only that the shares of types differ. Regarding q_h the trade-off is as in the full commitment case or for the no-audit contract, only that the weight on it is $1-\varrho$. Matters are slightly different for q_a, because this quantity is produced by both types. Still, the high-cost type is asked to produce q_a more frequently than the low-cost type, and the trade-off is determined by this margin. Because the latter margin is lower, the distortion on q_a is lower, hence $q_h^o > q_a > q_h$. Also, when concerned with q_a the rent left to the efficient type is reduced by P via the audit. The optimal communication mechanism when (2.16) is non-binding is given in the following proposition.

Proposition 2.3. *There exists a unique value $\underline{P}^m$, such that in the optimal audit mechanism under limited commitment* (2.16) *is not binding if and only if $P \leq \underline{P}^m$. The* *optimal communication mechanism then entails* $\varrho^* = 1,\ q_l^* = q_l^o$ *and* q_a^* *given by*

$$V'(q_a^*) = \theta_h + \frac{\phi P - c}{(1-\phi)P}\Delta\theta \tag{2.20}$$

The high-cost type earns no rent, whereas the low-cost type obtains a rent of $U_l^* = \Delta\theta q_a^* - P$.

The mechanism obtained in Proposition 2.3 shares many features with the corresponding contract under full commitment. In both settings, for low values of the deterrent P the rent to the efficient type cannot be reduced to zero. Auditing is then merely used to reduce the rent by as much as possible, which requires certain audits (after the respective type report).

At first sight, also quantities are altered when commitment is limited. After all, with full commitment the high-cost type produces q_h^{na} as long as rents are non-zero, whereas with limited commitment q_a is strictly higher. But this difference is a consequence of the need to incentivize the principal to actually conduct the required audits. Quantity q_a is artificially introduced to serve this purpose and because both agent's produce it, its value is determined differently from how q_h^{na} evolves. In the background, quantity q_h remains at the level q_h^{na}, but because the principal uses audits at the maximal frequency this quantity ultimately is never produced.

Further notice that $\underline{P}^m > \underline{P}^c$, i.e., it requires strictly larger deterrents to reduce all rents to zero under limited commitment. This immediately follows from the discussion above, because it is costly for the principal to introduce q_a.

We next consider larger deterrents for which (2.16) becomes binding and hence all rents are reduced to zero. As in the full commitment case, the principal now gradually reduces the distortions imposed on quantities. However, as outlined in the preceding paragraph, matters seem different from the outside. The following proposition characterizes the optimal communication mechanism for deterrents $P > \underline{P}^m$.

Proposition 2.4. *For deterrents* $P > \underline{P}^m$ *the low-cost types participation constraint is binding and hence* $U_h^* = U_l^* = 0$. *The optimal communication mechanism is given by either*

$$\varrho^* = 1,\ q_l^* = q_l^o \ \text{ and } \ q_a = P/\Delta\theta \tag{2.21}$$

or

$$\varrho^* < 1,\ q_l^* = q_l^o \ \text{ and } \ q_h^o > q_a > q_h > q_h^{na}. \tag{2.22}$$

For $P \geq \Delta\theta q_h^o$, the optimal mechanism always entails $\varrho^* < 1$.

The mechanism obtained from Proposition 2.4 shares many features with the full commitment contract. There may be a range where audits remain certain, i.e., all high-cost types are audited.[16] In this range, q_a raises linearly with P and actually corresponds to the quantity from the full commitment contract.[17] For large enough penalties audits get random, which means that not all high-cost types are audited. Under full commitment this simply means random audits after the agent reported to have high costs. With limited commitment the mechanism now uses all three quantities q_l, q_a and q_h. In particular this implies there are more contingencies than types.

Before focussing in greater detail on the contract with $\varrho^* < 1$, we briefly study implementability of the optimal mechanism as described in Propositions 2.3 and 2.4. When $\varrho^* = 0$ it is trivial that the optimal communication mechanism can be implemented via one round of face-to-face communication, i.e., using only modes of communication traditionally used in contract theory. This is also the case whenever $\varrho^* = 1$. To see this, notice that because (2.15) is binding and (2.16) is always satisfied, the low-cost type is actually indifferent between the outcomes t_l, q_l and no audit, and t_a, q_a and sure audit. Hence, this type is willing to randomize with reporting strategy as proposed by the communication mechanism. Furthermore, the high-cost type strictly prefers t_a, q_a over t_l, q_l.[18] Consequently, there exists a Perfect Bayesian Equilibrium of the game where the agent reports a type to the principal and the principal, after transfers were paid and quantities were produced, decides upon an audit. In this equilibrium the low-cost type randomizes between the two type reports and the high-cost type reports truthfully. The principal audits with certainty after the agent reported high costs.

The mechanism with $\varrho^* < 1$, however, cannot be implemented with one round of face-to-face communication. Assuming the randomization is executed by the agent, following Bester and Strausz (2001) there would exist an equivalent mechanism with only two reports for the agent. But this contradicts optimality of the above mechanism. Also, the randomization cannot be executed by the principal in form of a stochastic mechanism, because the implied knowledge of the agent's report when deciding upon

[16]Under limited commitment this implies that also a share of low-cost types is audited. Under full commitment the latter is not required.

[17]However, there need not be a range for the deterrent where the audit contracts under full and limited commitment coincide. In particular, audits can be already random under full commitment, while rents are still positive under limited commitment, i.e. when $\underline{P}^m > \overline{P}^c$.

[18]Also, this type is never affected by potential audits.

an audit. But the optimal communication mechanism requires that the principal does not know the agent's report after the randomization realized in production of q_a.

The following corollary formalizes the last statements.

Corollary 2.2. *The optimal communication mechanism can be implemented via one round of face-to-face communication if and only if $\varrho^* \in \{0, 1\}$.*

Let us now have a closer look at the random audit mechanism under limited commitment, i.e., the cases where $\varrho^* \in (0, 1)$. Each of the two type reports leads to a lottery. Overall there are three deterministic outcomes: t_l, q_l and no audits, t_a, q_a and sure audits, as well as t_h, q_h and no audits. Reporting to have low costs leads to a lottery over the first two outcomes, whereas reporting to have high costs leads to a lottery over the last two outcomes.

It is crucial that the randomization is not performed by the principal, but the mediator. In particular, when learning that the outcome t_a, q_a has been chosen, the principal is not supposed to know which type the agent reported. This distinguishes the optimal communication mechanism from a random contract, where the principal learns the reported type but nevertheless determines the outcome via a lottery. From the (truthful) type report to the mediator, the principal receives only the necessary information to persuade her to obey the recommendation. In case of t_a, q_a this implies, in particular, that the principal does not learn the (truthful) type report. Moreover, this cannot be achieved by letting the agent perform the randomization. There do not exist transfers t_l, t_a and t_h, such that each type of the agent is indifferent between the respective alternatives.

For a discussion of the underlying trade-offs it is helpful to solve (2.16) for ϱ:

$$\varrho = \frac{\Delta\theta q_h}{P - \Delta\theta(q_a - q_h)}. \tag{2.23}$$

From (2.23) it is immediate, that ϱ increases with q_a and q_h.

Increasing either q_a or q_h has two effects: First a *rent-effect* that can be best seen from looking at (2.19). Via the incentive constraint, an increase of the respective quantity increases the rent that has to be left to the low-cost type. The rent effect is stronger for q_h, because the low-cost type anyway produces q_a when $\varrho \in (0, 1)$. The second-effect we denote *audit-cost-effect*. It stems from the binding participation constraint of the low-cost type. All rents are reduced to zero via the use of audits. When a quantity is increased, also rents increase such that more audits are required to reduce those back to zero and more audits require a larger ϱ. But a larger ϱ implies a larger distortion via the 'mis-allocation', i.e., that each type produces q_a more often. What becomes crucial in comparing the rent- and the audit-cost-effect is, that the latter has the same (relative)[19] magnitude for both changes in q_a and q_h. Because the rent-effect differs for the respective quantities, this explains why q_a and q_h differ and why we have $q_a > q_h$.

With increasing P, the audit-cost-effect is weakened, because only a slight change in audit probability already has a severe impact on the agent's rent.[20] Therefore, quantities q_a and q_h increase with P and ϱ decreases. Also the difference between the rent-effects decreases with P, so the difference in the two quantities decreases with P.

Comparative statics when P is getting large are similar to those of the full commitment contract. In particular we also have $q_h < q_a < q_h^o$ for *all* finite P. A reduction away from q_h^o has no first-order effect on welfare from production, but allows for a reduction of audit costs (from 'mis-allocation'). This implies, that we have random audits for $P \geq \Delta\theta q_h^o$, both under full and limited commitment. As P approaches infinity, we further have that ϱ converges to zero and both q_a and q_h converge to q_h^o. Hence, in the limit, commitment plays no role, as expected.[21]

As in the case of full commitment, we cannot make a general statement about ϱ^* when $P \in (\underline{P}^m, \Delta\theta q_h^o)$. In general ϱ^* is not monotone on this range. In the following we work out an example, for which a unique threshold $\overline{P}^m$ exists, such that $\varrho^* < 1$ if and only if $P > \overline{P}^m$.

Example 2.1. *Let $V(q) = 2\sqrt{q}$ and the cost parameters be given by $\theta_l = 1$, resp., $\theta_h = 2$. Then $q_l^o = 1$ and $q_h^o = 1/4$. furthermore, assume $\phi = 1/2$, which yields $q_h^{na} = 1/9$ and the welfare from the no-audit contract equals $\mathcal{V}^{na} = 2/3$.*
Now consider audits and assume $c = 0.01$. The threshold-value P^* for profitable audits can be computed as $P^* \approx 0.0758$. Furthermore, we have $\underline{P}^m \approx 0.1241$. Consequently we have $\varrho^* = 1$ for all $P \in (P^*, \underline{P}^m]$. Lengthy calculations show that there exists a unique value $\overline{P}^m \approx 0.2387$ such that $\varrho^* = 1$ also for $P \in (\underline{P}^m, \overline{P}^m]$. On

[19]Relative in the following sense: Differentiating ϱ yields $\partial\varrho/\partial q_h = (1-\varrho)\varrho/q_h$ and $\partial\varrho/\partial q_a = \varrho^2/q_h$. When differentiating (2.19), the rent-effect has weight $1-\varrho$ for q_h, resp. ϱ for q_a. Factoring these weights out, the audit-cost-effect is the same, irrespective of the quantity.

[20]The relationship is in general not monotone. It depends on the shape of $V(\cdot)$, as already for the full commitment case. For large P this can be shown to hold in general, irrespective of the shape of $V(\cdot)$.

[21]The speed of convergence is different. For any finite P, the profits under full and limited commitment are different, but the gap closes, as P becomes large.

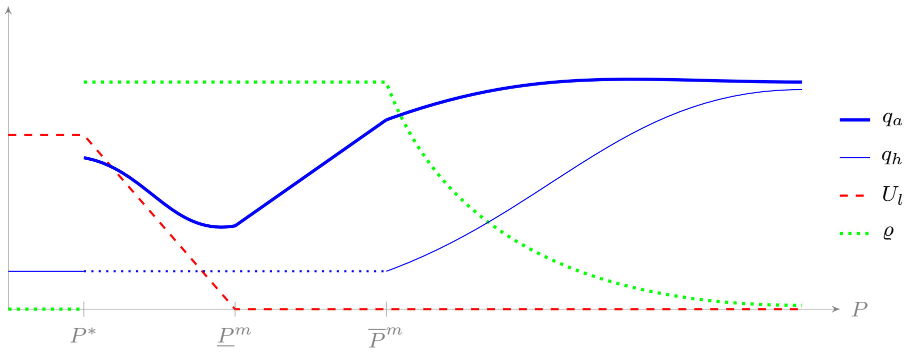

Figure 2.2: The mediated audit contract. Some characteristics as a function of P. Along the dotted blue line q_h is not produced.

this interval, we have $U_l = 0$ and hence three binding constraints. Only for $P > \overline{P}^m$, we have $\varrho^* < 1$. Notice further that $\overline{P}^m < \Delta\theta q_h^o$.

Figure 2.2 illustrates determinants of the communication mechanism as a function of the deterrent P in our example. The pattern for the efficient type's rents is as under full commitment, only that it requires a larger deterrent P to induce audits in the first place and also a larger penalty is required to reduce rents to zero. Similarly, the audit probability ϱ follows a similar pattern. Notice, however, that there are more audits under limited commitment because also some low-cost types are audited. The difference between optimal contracts under full and limited commitment can be best seen when looking at quantities. With limited commitment there is an upward jump at P^* when production switches from q_h^{na} to $\widehat{q}_a$. Furthermore, the optimal value $\widehat{q}_a$ declines with P as long as $P \le \underline{P}^m$. On the interval $(\underline{P}^m, \overline{P}^m)$ the pattern is similar as for the full commitment case. For $P > \overline{P}^m$ there are three quantities produced. Both q_a and q_h increase with P and tend to q_h^o as P tends to infinity.

2.6 Conclusion

This chapter analyzes the optimal coordination mechanism in a principal-agent framework when the principal cannot commit to an audit strategy. It is shown that to induce the principal to audit, the optimal mechanism provides an additional alternative that, if it is chosen induces the principal to audit with certainty. The ex-ante cost of an audit is determined by the welfare-effect of the new alternative. For large enough deterrents the optimal audit mechanism implements three distinct production schedules, despite there being only two types. Implementation via one single round of face-to-face communication is only possible, when the coordination mechanism uses only two production schedules.

We study a stylized two-type model. A natural question that arises is whether our results change when allowing for more types. The beneficial role of mediation carries over, but the analysis gets easily intractable. Already the full commitment case is messy to analyze with more than two types, see for instance Border and Sobel (1987). The problem lies in identifying the binding constraints. The possibility of an audit effectively transforms the one-dimensional screening problem into a multi-dimensional screening problem.

We leave open the question of whether there exist more elaborate face-to-face communication mechanisms, i.e., with multiple rounds of message exchange, that allow for implementing the optimal mechanism with $\varrho^* < 1$. Though this is relatively straightforward in the case of cheap talk, the presence of transfers makes the problem hard to analyze and we leave this issue for future research.

Our model is one of adverse selection, where the agent knows her type at the time of contracting. This differentiates our model from the literature on costly-state-verification, where contracting takes place *before* the agent obtains private information. Studying the role of mediation in a model with costly-stat-verification is therefore a next logical step.

Lastly, in the chapter we fully analyze the optimal contract in a setting of contracting with limited commitment. Our results provide new insights into both the structure of optimal communication mechanism and the analysis itself. Building on these insights may help studying mediation in different models, such as those of dynamic contracting with limited commitment (e.g., Laffont and Tirole (1988)), bilateral trade (e.g., Skreta (2006)) or auction design without commitment (e.g., Vartiainen (2013) and Skreta (2013)).

Appendix

2.A Proofs

2.A.1 Proofs of Section 2.4

Proof of Lemma 2.1. The proof is standard and therefore skipped. □

Proof of Lemma 2.2. First notice that we can set $\mathcal{P}_i(\theta_i) = 0$ for $i = l, h$ without loss of generality. To see this, suppose $\mathcal{P}_i(\theta_i) = P$. Then, setting $\widetilde{t}_i = t_i - \alpha_i P$ and $\widetilde{\mathcal{P}}_i(\theta_i) = 0$ keeps all constraints valid. However, the principal's profit is unchanged, because

$$\phi(t_i - \widetilde{t}_i) - \phi\alpha_i P_{ii} = \phi\alpha_i P - \phi\alpha_i P = 0. \tag{2.24}$$

In the following we solve a relaxed program, where we ignore the inefficient type's incentive constraint (2.10). Doing so, we can w.l.o.g. set $\mathcal{P}_l = \mathcal{P}^0$, because penalties to the high-cost type do not affect any of the three remaining constraints. Furthermore, we can set $\mathcal{P}_h = \mathcal{P}^l$.

It is then straightforward that (2.8) and (2.9) are binding. Substituting transfers from these constraint into the principal's objective yields

$$\phi\big(V(q_l) - \theta_l q_l - \Delta\theta q_h + \alpha P\big) + (1-\phi)\big(V(q_h) - \theta_h q_h - \alpha c\big). \tag{2.25}$$

Hence, the principal maximizes (2.25) with respect to (2.7). First maximize (2.25) while ignoring (2.7). This yields $q_l = q_l^o$ and $q_h = q_h^{na}$. As long as $P \leq \underline{P}^c := \Delta\theta q_h^{na}$ this solution does indeed satisfy (2.7). Otherwise, there exists a unique $\alpha(P) \in (0, 1)$ such that (2.7) is satisfied for $\alpha \leq \alpha(P)$.

Now add the efficient type's participation constraint to the problem, for a given value of α the Lagrangian reads as

$$\begin{aligned}\mathcal{L}(\alpha, q_l, q_h, \lambda) = \phi\big(V(q_l) - \theta_l q_l - \Delta\theta q_h + \alpha P\big) \\ + (1-\phi)\big(V(q_h) - \theta_h q_h - \alpha c\big) + \lambda(\Delta\theta q_h - \alpha P).\end{aligned} \tag{2.26}$$

Maximization over q_l, q_h and λ yields the first-order conditions

$$\begin{aligned} V'(q_l) &= \theta_l \\ V'(q_h) &= \theta_h + \frac{\phi-\lambda}{1-\phi}\Delta\theta \\ \Delta\theta q_h - \alpha P &= 0. \end{aligned}$$

Define $\mathcal{V}^c(\alpha) := \max_{q_l, q_a, \lambda} \mathcal{L}(\alpha, q_l, q_h, \lambda)$. By the envelope-theorem we have

$$\frac{\partial \mathcal{V}^c}{\partial \alpha} = \phi P - (1-\phi)c - \lambda P. \tag{2.27}$$

Because $\alpha(P)$, as defined above, is strictly positive for all P we have $\partial\mathcal{V}^c/\partial\alpha = \phi P - (1-\phi)c$ for $\alpha \in [0, \alpha(P)]$ for all P. Furthermore, by the implicit function theorem we have $\partial\lambda/\partial\alpha = -P(1-\phi)V''(q_h)/((\Delta\theta)^2) > 0$. Hence

$$\frac{\partial^2 \mathcal{V}^c}{\partial \alpha^2} = -\frac{\partial\lambda}{\partial\alpha}P < 0 \tag{2.28}$$

for $\alpha \in (\alpha(P), 1]$.

This proves that we have three cases for the optimal value of α^c:

- $\alpha^c = 0$, which applies whenever $\phi P \leq (1-\phi)c$.
- $\alpha^c = 1$, which applies whenever $(\partial\mathcal{V}^c)/(\partial\alpha) \geq 0$ at $\alpha = 1$.
- $\alpha^c = (\alpha(P), 1)$, which applies whenever $\phi P > (1-\phi)c$ and $(\partial\mathcal{V}^c)/(\partial\alpha) < 0$ at $\alpha = 1$.

For the remainder assume $\phi P > (1-\phi)c$. For $P \leq \underline{P}^c$ we have that $\mathcal{L}^c(\alpha)$ is linear in α and consequently maximized at $\alpha = 1$. For $P > \underline{P}^c$ notice that $\partial\mathcal{V}^c/\partial\alpha = 0$ if and only if $\phi - \lambda = (1-\phi)c/P$ and therefore

$$V'(q_h) = \theta_h + \frac{c}{P}\Delta\theta > \theta_h. \tag{2.29}$$

Consequently, for $\alpha^c \in (0,1)$ we have that q_h^c satisfies the above condition. Notice that for $P = \Delta\theta q_h^{na}$ we have $V'(P/\Delta\theta) > \theta_h + \frac{c}{P}\Delta\theta$ and consequently the solution has $\alpha^c = 1$, which by continuity holds also on some interval $[\Delta\theta q_h^{na}, \Delta\theta q_h^{na} + \varepsilon)$. Furthermore (2.29) implies that $\alpha^c < 1$ for all $P \geq \Delta\theta q_h^o$. Hence, there exists a $\overline{P}^c \in (\underline{P}^c, \Delta\theta q_h^o)$ such that $\alpha < 1$ for all $P > \overline{P}^c$. If on the other hand $\alpha^c = 1$ and (2.7) binding, we must have $q_h = P/\Delta\theta$.

Finally, it is straightforward to verify that (IC$_h$) is indeed satisfied: We have $t_l \leq \theta_l q_l^o + \Delta\theta q_h^{na}$ and therefore $t_l - \theta_h q_l^o \leq -\Delta\theta(q_l^o - q_h^{na}) < 0$. □

Proof of Corollary 2.1. Taking the derivative of $\partial\mathcal{V}^c/\partial\alpha$ with respect to P yields

$$\frac{\partial^2\mathcal{V}^c}{\partial\alpha\partial P} = (\phi - \lambda) - \frac{\partial\lambda}{\partial P}P. \tag{2.30}$$

From the implicit function theorem we have $\partial\lambda/\partial P = -\alpha V''(q_h)(1-\phi)/((\Delta\theta)^2)$, consequently

$$\begin{aligned}
\frac{\partial^2\mathcal{V}^c}{\partial\alpha\partial P} &= \frac{1-\phi}{\Delta\theta}\left(\frac{\phi-\lambda}{1-\phi}\Delta\theta + V''(q_h)q_h\right) \\
&\leq \frac{1-\phi}{\Delta\theta}\left(\frac{\phi-\lambda}{1-\phi}\Delta\theta - V'(q_h)\frac{\phi\Delta\theta}{(1-\phi)\theta_h + \phi\Delta\theta}\right) \\
&= \frac{1-\phi}{\Delta\theta}\left(\frac{\phi-\lambda}{1-\phi}\Delta\theta - (\theta_h + \frac{\phi-\lambda}{1-\phi}\Delta\theta)\frac{\phi\Delta\theta}{(1-\phi)\theta_h + \phi\Delta\theta}\right) \\
&= (\phi-\lambda)\frac{(1-\phi)\theta_h}{(1-\phi)\theta_h + \phi\Delta\theta} - \phi\frac{(1-\phi)\theta_h}{(1-\phi)\theta_h + \phi\Delta\theta} \\
&= -\lambda\frac{(1-\phi)\theta_h}{(1-\phi)\theta_h + \phi\Delta\theta} < 0.
\end{aligned}$$

We know already that $\alpha_h^c < 1$ for $P \geq \Delta\theta q_h^o$. The above proves, that there exists a unique value $\overline{P}^c$ such that $\alpha_h^c < 1$ if and only if $P > \overline{P}^c$. Moreover, we have $\partial\alpha_h^c/\partial P < 0$.

Considering q_h^c, we have that $q_h^c \equiv q_h^{na}$ for all $P \leq \underline{P}^c$. For $P \in (\underline{P}^c, \overline{P}^c]$ we have $q_h^c = P/\Delta\theta$ which increases with P. Lastly, for $P > \overline{P}^c$ we have $\partial q_h^c/\partial P = -(\Delta\theta c)/(P^2 V''(q_h^c)) > 0$. □

2.A.2 Proofs of Section 2.5

Proof of Proposition 2.1. To simplify notation, let $d = (r, q, t, \mathcal{P}(\cdot)) \in \{a, na\} \times \mathbb{R}_+ \times \mathbb{R} \times \{\mathcal{P}^0, \mathcal{P}^l, \mathcal{P}^h, \mathcal{P}^{lh}\}$ be shorthand for an allocation determined by the mediator. For a given $\boldsymbol{\pi}$ define $\mathrm{supp}\{\boldsymbol{\pi}\} := \mathrm{supp}\{\pi_l\} \cup \mathrm{supp}\{\pi_h\}$ the support of $\boldsymbol{\pi}$, i.e. the set of allocations d that are chosen with strictly positive probability for *some* type report. A mechanism $\boldsymbol{\pi}$ is *incentive compatible* if

$$\sum_{d \in \mathrm{supp}\{\pi_i\}} \left[t - \theta_i q - \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_i)\right] \pi_i(d) \ \geq \sum_{d \in \mathrm{supp}\{\pi_j\}} \left[t - \theta_i q - \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_i)\right] \pi_j(d), \tag{IC$_i$}$$

holds for all $i \neq j \in \{l, h\}$, i.e. each type of the agent prefers truthful reporting; and

$$(-1)^{\mathbb{1}_{\{r=na\}}} \left\{ \frac{\phi \pi_l(d)}{\phi \pi_l(d) + (1-\phi)\pi_h(d)} \mathcal{P}(\theta_l) + \frac{(1-\phi)\pi_h(d)}{\phi \pi_l(d) + (1-\phi)\pi_h(d)} \mathcal{P}(\theta_h) - c \right\} \geq 0, \tag{OC$_d$}$$

for all $d \in \mathrm{supp}\{\boldsymbol{\pi}\}$, i.e. the principal obediently follows the recommended action.

The mechanism is *individually rational* if

$$\sum_{d \in \mathrm{supp}\{\pi_i\}} \left[t - \theta_i q - \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_i)\right] \pi_i(d) \geq 0, \tag{IR$_i$}$$

for $i = l, h$ - each type receives at least the outside option of zero. Finally, $\boldsymbol{\pi}$ is *feasible*, whenever

$$\pi_i(d) \geq 0 \quad \forall i, d; \qquad \sum_{d \in \mathrm{supp}\{\pi_i\}} \pi_i(d) = 1, \tag{FC}$$

that is whenever each π_i is a proper probability distribution.

The principal's expected profit from applying the incentive compatible, individually rational and feasible mechanism $\boldsymbol{\pi}$ is

$$\begin{aligned}\mathcal{V}(\boldsymbol{\pi}) = \quad & \phi \sum_{d \in \operatorname{supp}\{\pi_l\}} \Big[V(q) - t + \mathbb{1}_{\{r=a\}}(\mathcal{P}(\theta_l) - c) \Big] \pi_l(d) \\ & + (1-\phi) \sum_{d \in \operatorname{supp}\{\pi_h\}} \Big[V(q) - t + \mathbb{1}_{\{r=a\}}(\mathcal{P}(\theta_h) - c) \Big] \pi_h(d).\end{aligned} \tag{2.31}$$

This allows us to state the principal's problem of finding the optimal coordination mechanism:

$$\max_{\boldsymbol{\pi}} \mathcal{V}(\boldsymbol{\pi}\} \quad \text{s.t. } (\text{IR}_i), (\text{IC}_i), (\text{OC}_d), (\text{FC}) \quad \text{for } i = l, h \text{ and } d \in \operatorname{supp}\{\boldsymbol{\pi}\}. \tag{$\mathcal{P}$}$$

Instead of solving problem ($\mathcal{P}$) we shall set up an auxiliary problem that is easier to solve and at the end argue that the solution of the auxiliary problem also solves problem ($\mathcal{P}$).

First we dispose the irrelevant obedience constraints. As stated in (OC_d), we impose obedience constraints for both recommended actions. However, if the recommended action is 'no audit', obedience is easily guaranteed if we set $\mathcal{P} \equiv \mathcal{P}^0$. Because the principal observes the penalty scheme, this change does not affect any of the other constraints, hence we can focus on obedience constraints where the recommended action is audit:

$$\begin{aligned}\Bigg\{ & \frac{\phi \pi_l(d)}{\phi \pi_l(d) + (1-\phi)\pi_h(d)} \mathcal{P}(\theta_l) \\ & + \frac{(1-\phi)\pi_h(d)}{\phi \pi_l(d) + (1-\phi)\pi_h(d)} \mathcal{P}(\theta_h) - c \Bigg\} \geq 0,\end{aligned} \tag{OC'_d}$$

for all $d \in \operatorname{supp}\{\boldsymbol{\pi}\}$ and $r = a$.

Next, we proceed as in the full commitment case and guess that (IC_h) is not binding, which we will verify at the end.[22]

Finally, a simplification that helps us to substantially reduce the complexity of the following analysis is to focus on *type-dependent* transfers. In particular, we substitute $\sum_d t\, \pi_i(d)$ by T_i. This allows for more flexibility in the principal's problem

[22] The problem here and in the full commitment case is that the screening problem is essentially multi-dimensional through the impact of the penalty schemes. Potential penalties for the inefficient type may render (IC_h) binding. By assuming the constraint to be slack, we can easily rule out these penalties, which ultimately helps justifying the assumption in the first place.

and turns out to be analytically more tractable. Notice however that this constitutes a purely theoretical simplification, because the principal will observe the transfer paid to the agent. Hence, at the end of our analysis we shall point out how to re-transform the type-dependent transfers into allocation-specific transfers. The agent's individual rationality constraints now read as

$$T_i - \sum_{d \in \text{supp}\{\pi_i\}} \Big[\theta_i q + \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_i)\Big] \pi_i(d) \ \geq 0, \tag{IR$'_i$}$$

for $i = l, h$, and the efficient type's incentive compatibility constraint is

$$\begin{aligned} T_l - \sum_{d \in \text{supp}\{\pi_l\}} \Big[\theta_l q + \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_l)\Big] \pi_l(d) \ \geq \\ T_h - \sum_{d \in \text{supp}\{\pi_h\}} \Big[\theta_l q + \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_l)\Big] \pi_h(d). \end{aligned} \tag{IC$'_l$}$$

The principal's profit can now be stated as

$$\begin{aligned} \mathcal{V}(\boldsymbol{\pi}) = \quad & \phi \sum_{d \in \text{supp}\{\pi_l\}} \Big[V(q) + \mathbb{1}_{\{r=a\}} (\mathcal{P}(\theta_l) - c)\Big] \pi_l(d) \\ & + (1-\phi) \sum_{d \in \text{supp}\{\pi_h\}} \Big[V(q) + \mathbb{1}_{\{r=a\}} (\mathcal{P}(\theta_h) - c)\Big] \pi_h(d) \\ & - \phi T_l - (1-\phi) T_h. \end{aligned} \tag{2.32}$$

The auxiliary problem we solve in the following is

$$\max_{\boldsymbol{\pi}} \mathcal{V}(\boldsymbol{\pi}\} \quad \text{s.t. (IR}'_i), (\text{IC}'_l), (\text{OC}'_d), (\text{FC}) \quad \text{for } i = l, h \text{ and } d \in \text{supp}\{\boldsymbol{\pi}\}. \tag{$\mathcal{P}'$}$$

Notice that we abuse notation here: After moving towards type-dependent transfers, an allocation d consists only of a quantity q, a recommendation r and a penalty scheme $\mathcal{P}(\cdot)$. Consequently, randomizations are understood to be over these variables and the support changes accordingly.

The proof is now a sequence of intermediate results. The following two Lemmas rule out penalties for the inefficient type.

Lemma 2.3. *It is never optimal to have strictly positive mass on the penalty scheme $\mathcal{P}^{lh}$.*

Proof. Suppose $\boldsymbol{\pi}$ satisfies all constraints of problem ($\mathcal{P}'$) and there exists q such that $\pi_i(q, \mathcal{P}^{lh}, a) > 0$ for some $i \in \{l, h\}$.[23] Consider $\widetilde{\boldsymbol{\pi}}$, where T_i is replaced by $\widetilde{T}_i = T_i - \pi_i(q, \mathcal{P}^{lh}, a)P$. Furthermore, set $\widetilde{\pi}_i(q, \mathcal{P}^{lh}, a) = 0$, and $\widetilde{\pi}_i(q, \mathcal{P}^0, na) = \pi_i(q, \mathcal{P}^0, na) + \pi_i(q, \mathcal{P}^{lh}, a)$. In all other respects $\widetilde{\boldsymbol{\pi}}$ coincides with $\boldsymbol{\pi}$. It is straightforward to verify that $\widetilde{\boldsymbol{\pi}}$ satisfies all constraints of problem ($\mathcal{P}'$) and yields expected profit $\mathcal{V}(\widetilde{\boldsymbol{\pi}}) = \mathcal{V}(\boldsymbol{\pi}) + (\pi_l(q, \mathcal{P}^{lh}, a) + \pi_h(q, \mathcal{P}^{lh}, a))c > \mathcal{V}(\boldsymbol{\pi})$. □

Lemma 2.4. *Without loss of generality we can assume that the optimal mechanism has no strictly positive mass on the penalty scheme $\mathcal{P}^h$.*

Proof. Suppose $\boldsymbol{\pi}$ satisfies all constraints of problem ($\mathcal{P}'$) and there exists q such that $\pi_h(q, \mathcal{P}^h, a) > 0$. Consider $\widetilde{\boldsymbol{\pi}}$, where we set $\widetilde{\pi}_i(q, \mathcal{P}^h, a) = 0$ and $\widetilde{\pi}_i(q, \mathcal{P}^0, na) = \pi_i(q, \mathcal{P}^0, na) + \pi_i(q, \mathcal{P}^h, a)$. Further set the transfer to the inefficient type as $\widetilde{T}_h = T_h - \pi_h(q, \mathcal{P}^h, a)P$. Clearly, $\widetilde{\boldsymbol{\pi}}$ satisfies all constraints of problem ($\mathcal{P}'$) and the for the respective profits we have $\mathcal{V}(\widetilde{\boldsymbol{\pi}}) \geq \mathcal{V}(\boldsymbol{\pi})$. □

Lemmas 2.3 and 2.4 imply that a solution to problem ($\mathcal{P}'$), without loss of generality, puts strictly positive mass only on outcomes involving the penalty schemes $\mathcal{P}^0$ and $\mathcal{P}^l$. Furthermore, the penalty scheme $\mathcal{P}^l$ always accompanies the recommendation to audit, whereas $\mathcal{P}^0$ always implies $r = na$. The following Lemma argues that the principal is kept indifferent whether the obey the mediators recommendation to audit.

Lemma 2.5. *If the principal is recommended to audit, then she is just indifferent whether to obey this recommendation.*

Proof. Fix some outcome $(q, \mathcal{P}^l, a)$ such that $\pi_l(q, \mathcal{P}^l, a) > 0$. Assume by contradiction, that the principal strictly prefers to obey the mediator's recommendation. Consider the alternative mechanism $\widetilde{\boldsymbol{\pi}}$ which coincides with $\boldsymbol{\pi}$, apart from

$$\widetilde{\pi}_l(q, \mathcal{P}^l, a) = \pi_l(q, \mathcal{P}^l, a) - \varepsilon, \tag{2.33}$$

and

$$\widetilde{\pi}_l(q, \mathcal{P}^0, a) = \pi_l(q, \mathcal{P}^0, a) + \varepsilon, \tag{2.34}$$

for some small $\varepsilon > 0$. Lastly, set $\widetilde{T}_l = T_l - \varepsilon P$. As long as ε is small enough, the principal still obediently follows the mediator's recommendations. Furthermore, (IR$_h$)

[23] As argued in the paragraph preceeding OC$'_d$ we assume that $\mathcal{P} \equiv \mathcal{P}^0$ whenever no audit is recommended.

is unaffected and the adjusted transfer just keeps (IC$_l$) and (IR$_l$) valid. However, we have $\mathcal{V}(\widetilde{\boldsymbol{\pi}}) - \mathcal{V}(\boldsymbol{\pi}) = \varepsilon P - \varepsilon(P - c) = \varepsilon c > 0$. $\square$

By what we have shown so far, there can be three distinct outcome types: a quantity q produced by the l-type and no audit recommended, a quantity q produced by the h-type and no audit recommended, and a quantity q produced by either type with the recommendation to audit. We shall prove in the remainder that each invoked outcome is unique with respect to the quantity.

At this point, notice that both (IR$_h$) and (IC$_l$) must be binding: If the former was not binding, the principal could increase profits by lowering T_h without violating any other constraint. The latter must be binding, because otherwise the first-best was a solution to the principal's problem. But this clearly violates (IC$_l$).

Because by Lemmas 2.3 and 2.4 the inefficient type is never penalized, we get from (IR$_h$) that

$$T_h = \sum_{d \in \text{supp}\{\pi_h\}} \theta_h q \, \pi_h(d). \tag{2.35}$$

The binding constraint (IC$_l$) can then be solved for T_l as follows:

$$\begin{aligned} T_l = & \sum_{d \in \text{supp}\{\pi_l\}} \left[\theta_l q + \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_l) \right] \pi_l(d) \\ & + \sum_{d \in \text{supp}\{\pi_h\}} \left[\Delta\theta q - \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_l) \right] \pi_h(d). \end{aligned} \tag{2.36}$$

Substituting these transfers into the principal's objective yields

$$\begin{aligned} \mathcal{V}(\boldsymbol{\pi}) = \phi & \sum_{d \in \text{supp}\{\pi_l\}} (V(q) - \theta_l q) \, \pi_l(d) \\ & - \phi \sum_{d \in \text{supp}\{\pi_h\}} \left[\Delta\theta q - \mathbb{1}_{\{r=a\}} \mathcal{P}(\theta_l) \right] \pi_h(d) \\ & + (1 - \phi) \sum_{d \in \text{supp}\{\pi_h\}} (V(q) - \theta_h q) \, \pi_h(d). \end{aligned} \tag{2.37}$$

Lemma 2.6. *Suppose $\boldsymbol{\pi}$ is a solution to problem ($\mathcal{P}'$) and $\pi_l(q, \mathcal{P}^0, na) > 0$. Then $q = q_l^o$.*

Proof. Follows immediately from (2.37) and noticing that $V(q) - \theta_l q$ has the unique maximizer q_l^o. Notice that any modification of a quantity for which $\pi_l(q, \mathcal{P}^0, na) > 0$

that is accompanied by an adjustment of T_l to keep the left-hand side of (IC$_l$) unchanged, does also not affect (IR$_l$). Hence, focussing solely on (2.37) is indeed sufficient. $\square$

Lemma 2.7. *Suppose $\boldsymbol{\pi}$ is a solution to problem ($\mathcal{P}'$). There exists at most one q such that $\pi_h(q, \mathcal{P}^0, na) > 0$.*

Proof. Suppose the contrary, i.e., there exists $d = (q, \mathcal{P}^0, na)$ and $\widehat{d} = (\widehat{q}, \mathcal{P}^0, na)$ with $q \neq \widehat{q}$ and both $\pi_h(d) > 0$ and $\pi_h(\widehat{d}) > 0$. Set

$$\widetilde{q} := \frac{q\pi_h(d) + \widehat{q}\pi_h(\widehat{d})}{\pi_h(d) + \pi_h(\widehat{d})}. \tag{2.38}$$

The value $\widetilde{q}$ is the probability-weighted average of the two quantities q and $\widehat{q}$. Further consider $\widetilde{\boldsymbol{\pi}}$, where we set $\widetilde{\pi}_h(\widetilde{q}, \mathcal{P}^0, na) = \pi_h(d) + \pi_h(\widehat{d})$. All other entries of $\widetilde{\boldsymbol{\pi}}$ coincide with those of $\boldsymbol{\pi}$. The agent's constraints are all linear in quantities, that is why the validity of all the agent's constraints for the mechanism $\widetilde{\boldsymbol{\pi}}$, follows from validity for the mechanism $\boldsymbol{\pi}$. Because the function $(1-\phi)(V(q) - \theta_h q) - \phi\Delta\theta q$ is strictly concave, we have $\mathcal{V}(\widetilde{\boldsymbol{\pi}}) > \mathcal{V}(\boldsymbol{\pi})$. $\square$

Lemma 2.8. *Suppose $\boldsymbol{\pi}$ is a solution to problem ($\mathcal{P}'$). There exists at most one q such that $\pi_l(q, \mathcal{P}^l, a) > 0$.*

Proof. Suppose the contrary, i.e., there exists $d = (q, \mathcal{P}^l, a)$ and $\widehat{d} = (\widehat{q}, \mathcal{P}^l, na)$ with $q \neq \widehat{q}$ and both $\pi_l(d) > 0$ and $\pi_l(\widehat{d}) > 0$. Then also $\pi_h(d) > 0$ and $\pi_h(\widehat{d}) > 0$ by Lemma 2.5. Set

$$\widetilde{q} := \frac{q\pi_l(d) + \widehat{q}\pi_l(\widehat{d})}{\pi_l(d) + \pi_l(\widehat{d})}. \tag{2.39}$$

Furthermore consider $\widetilde{\boldsymbol{\pi}}$, where we set $\widetilde{\pi}_l(\widetilde{q}, \mathcal{P}^l, a) = \pi_l(d) + \pi_l(\widehat{d})$, as well as $\widetilde{\pi}_h(\widetilde{q}, \mathcal{P}^l, a) = \pi_h(d) + \pi_h(\widehat{d})$. Also, set $\widetilde{\pi}_i(d) = \widetilde{\pi}_i(\widehat{d}) = 0$. All other entries of $\widetilde{\boldsymbol{\pi}}$ coincide with those of $\boldsymbol{\pi}$. From the principal's indifference whether to follow the recommendation to audit we get

$$\pi_h(d) = \frac{\phi(P-c)}{(1-\phi)c}\pi_l(d), \qquad \pi_h(\widehat{d}) = \frac{\phi(P-c)}{(1-\phi)c}\pi_l(\widehat{d}), \tag{2.40}$$

and therefore

$$\begin{aligned}\widetilde{\pi}_h(\widetilde{q},\mathcal{P}^l,a)\ \widetilde{q} =& \big(\pi_h(d)+\pi_h(\widehat{d})\big)\ \widetilde{q}\\ =& \frac{\phi(P-c)}{(1-\phi)c}\big(q\pi_l(d)+\widehat{q}\pi_l(\widehat{d})\big)\\ =& q\pi_h(d)+\widehat{q}\pi(\widehat{d}).\end{aligned}$$

Thus, as in the proof of Lemma 2.7 validity of the agent's constraints for mechanism $\boldsymbol{\pi}$ implies validity of the same constraints for mechanism $\widetilde{\boldsymbol{\pi}}$. Strict concavity of the functions $V(q)-\theta_l q$, resp., $(1-\phi)(V(q)-\theta_h q)-\phi\Delta\theta q$ then yields $\mathcal{V}(\widetilde{\boldsymbol{\pi}})>\mathcal{V}(\boldsymbol{\pi})$. □

Trivially, Lemma 2.8 also implies there exists at most one value q s.t. $\pi_h(q,\mathcal{P}^l,a)>0$, and, hence, we have shown that in any solution to problem ($\mathcal{P}'$) there are at most three distinct quantities. Denote them q_l, q_a and q_h, as prescribed in Proposition 2.1.

What remains to show is, that these insights extend to the solution of problem ($\mathcal{P}$). To show this, let $\boldsymbol{\pi}$ be a solution to problem ($\mathcal{P}'$). Setting $t_a=\theta_h q_a$ and $t_h=\theta_h q_h$, it is straightforward to verify that (IR$_h$) and the right-hand side of (IC$_l$) are kept unchanged. Solving (IC$_l$) for t_l yields validity of both (IC$_l$) and (IR$_l$). Obviously, all obedience constraints are satisfied. What remains to be shown is validity of (IC$_h$). For this we have to anticipate some of the results obtained in Propositions 2.2 - 2.4. These Propositions solve also for optimal quantities and probabilities of the solution to problem $\mathcal{P}'$. There are three cases:

(1) (no-audit contract)
$\pi_l(q_a,\mathcal{P}^l,a)=\pi_h(q_a,\mathcal{P}^l,a)=0$. Then $q_l=q_l^o$ and $q_h=q_h^{na}$ and the h-type clearly receives a strictly negative payoff from misreporting.

(2) (sure-audit contract, see Propositions 2.3 and 2.4)
$\pi_h(q_a,\mathcal{P}^l,a)=1$ which implies $\pi_l(q_a,\mathcal{P}^l,a)=((1-\phi)c)/(\phi(P-c))$. The high-cost type's utility from misreporting can be simplified to

$$-\frac{\phi P-c}{\phi(P-c)}\cdot\Delta\theta(q_l-q_a+P), \tag{2.41}$$

which is negative, because as we shall argue in the main text we must have $\phi P>c$ and $q_l>q_a$ in this case.

(3) (random-audit contract, see Proposition 2.4)
$\pi_h(q_a, \mathcal{P}^l, a) = \varrho \in (0,1)$ and $\pi_l(q_a, \mathcal{P}^l, a) = \varrho((1-\phi)c)/(\phi(P-c))$. Using the binding constraint (IR$_l$), the payoff of type θ_h from misreporting can be simplified to

$$\left(1 - \varrho\frac{(1-\phi)c}{\phi(P-c)}\right)\left(-\Delta\theta q_l - \varrho\frac{(1-\phi)c}{\phi(P-c)}(P - \Delta\theta q_a)\right) < 0, \tag{2.42}$$

where the strict inequality follows from $P > \Delta\theta q_a$ and $\phi P > c$.

This completes the proof of Proposition 2.1. □

Proof of Proposition 2.2. First ignore (IR$_l$). For a given $\varrho \in (0,1)$ differentiate (2.17) with respect to quantities. This yields $q_l = q_l^o, q_h = q_h^{na}$, and $q_a = \widehat{q}_a$, where

$$V'(\widehat{q}_a) = \theta_h + \frac{\phi P - c}{(1-\phi)P}\Delta\theta. \tag{2.43}$$

Now define $\underline{P}^m$ as the solution of

$$V'\left(\frac{P}{\Delta\theta}\right) = \theta_h + \frac{\phi P - c}{(1-\phi)P}\Delta\theta. \tag{2.44}$$

The value $\underline{P}^m$ exists, because for $P \to \infty$ we have $(\phi P - c)/((1-\phi)P) \to \phi/(1-\phi) > 0$, while $V(P/\Delta\theta) \to 0$, and for $P \to 0$ the right-hand side of (2.44) gets negative, whereas the left-hand side is always positive. It is unique, because the left-hand side of (2.44) strictly decreases with P, whereas the right-hand side strictly increases.

By the definition of $\widehat{q}_a$ we have $\Delta\theta\widehat{q}_a \geq P$ if and only if $P \leq \underline{P}^m$. Provided (IR$_h$) and (IC$_l$) are binding, (IR$_l$) rewrites as

$$\varrho\Delta\theta q_a + (1-\varrho)\Delta\theta q_h - \varrho P \geq 0. \tag{2.45}$$

Using the values for quantities derived above, this condition is satisfied for all $\varrho \in (0,1)$, whenever $\Delta\theta\widehat{q}_a - P \geq 0$, i.e. whenever $P \leq \underline{P}^m$. When $P > \underline{P}^m$, there exists a unique $\varrho(P) \in (0,1)$ such that $\varrho\Delta\theta q_a + (1-\varrho)\Delta\theta q_h - \varrho P \geq 0$ if and only if $\varrho \leq \varrho(P)$.

Next maximize (2.17) for a given $\varrho > \varrho(P)$ together with the now binding constraint (IR$_l$). Let λ denote the Lagrange-multiplier on (IR$_l$), the Lagrangian reads as

$$\begin{aligned}\mathcal{L}(q_l, q_a, q_h, \lambda) = \quad & \phi\big(V(q_l) - \theta_l q_l - \Delta\theta q_h\big) + (1-\phi)\big(V(q_h) - \theta_h q_h\big) \\ & - \varrho\bigg\{\frac{(1-\phi)c}{P-c}\big(V(q_l) - \theta_l q_l - V(q_a) + \theta_l q_a\big) \\ & - (1-\phi)\big(V(q_a) - \theta_h q_a - V(q_h) + \theta_h q_h\big) \\ & + \phi\Delta\theta(q_a - q_h) - \frac{\phi P - c}{P-c}P\bigg\} \\ & + \lambda\big(\varrho\Delta\theta q_a + (1-\varrho)\Delta\theta q_h - \varrho P\big).\end{aligned} \tag{2.46}$$

First order conditions for quantities and λ are

$$q_l: \quad V'(q_l) = \theta_l, \tag{2.47}$$

$$q_a: \quad V'(q_a) = \theta_h + \frac{\phi P - c - \lambda(P-c)}{(1-\phi)P}\Delta\theta, \tag{2.48}$$

$$q_h: \quad V'(q_h) = \theta_h + \frac{\phi - \lambda}{1-\phi}\Delta\theta, \tag{2.49}$$

$$\lambda: \quad \varrho\Delta\theta q_a + (1-\varrho)\Delta\theta q_h - \varrho P. \tag{2.50}$$

Thus, once more we have $q_l = q_l^o$, whereas $q_a > \widehat{q}_a$ and $q_h > q_h^{na}$ because $\lambda > 0$. Define

$$\mathcal{V}(\varrho) := \max_{q_l, q_a, q_h, \lambda} \mathcal{L}(q_l, q_a, q_h, \lambda) \tag{2.51}$$

Clearly, this function is continuous on the whole interval $[0, 1]$. Consider the following cases:

- (I) $P \leq \underline{P}^m$. Then $\mathcal{V}(\varrho)$ is linear in ϱ, because all quantities do not depend on ϱ and we have $\lambda = 0$. Furthermore,

$$\begin{aligned}\frac{\partial\mathcal{V}}{\partial\varrho} = \; & -\bigg\{\frac{(1-\phi)c}{P-c}\big(V(q_l^o) - \theta_l q_l^o - V(\widehat{q}_a) + \theta_l\widehat{q}_a\big) \\ & - (1-\phi)\big(V(\widehat{q}_a) - \theta_h\widehat{q}_a - V(q_h^{na}) + \theta_h q_h^{na}\big) \\ & + \phi\Delta\theta(\widehat{q}_a - q_h^{na}) - \frac{\phi P - c}{P-c}P\bigg\}.\end{aligned} \tag{2.52}$$

Hence, $\partial\mathcal{V}/\partial\varrho > 0$ if and only if $P > P^*$, as defined in (2.18).

- For $P > \underline{P}^m$ we have that (2.52) holds for $\varrho \leq \varrho(P)$. For $\varrho > \varrho(P)$ we have, after applying the envelope-theorem

$$\begin{aligned}\frac{\partial \mathcal{V}}{\partial \varrho} = & -\Bigg\{\frac{(1-\phi)c}{P-c}\big(V(q_l)-\theta_l q_l - V(q_a)+\theta_l q_a\big) \\ & -(1-\phi)\big(V(q_a)-\theta_h q_a - V(q_h)+\theta_h q_h\big) \\ & +\phi\Delta\theta(q_a-q_h)-\frac{\phi P-c}{P-c}P\Bigg\} \\ & +\lambda\big(\Delta\theta(q_a-q_h)-P\big).\end{aligned} \tag{2.53}$$

Differentiating once again yields

$$\begin{aligned}\frac{\partial^2\mathcal{V}}{\partial\varrho^2} &= \frac{(1-\phi)c}{P-c}(V'(q_a)-\theta_l)\frac{\partial q_a}{\partial\varrho}-\phi\Delta\theta\left(\frac{\partial q_a}{\partial\varrho}-\frac{\partial q_h}{\partial\varrho}\right) \\ &\quad +(1-\phi)(V'(q_a)-\theta_h)\frac{\partial q_a}{\partial\varrho}-(1-\phi)(V'(q_h)-\theta_h)\frac{\partial q_h}{\partial\varrho} \\ &\quad +\big(\Delta\theta(q_a-q_h)-P\big)\frac{\partial\lambda}{\partial\varrho}+\lambda\Delta\theta\left(\frac{\partial q_a}{\partial\varrho}-\frac{\partial q_h}{\partial\varrho}\right) \\ &= (1-\lambda)\frac{c}{P}\Delta\theta\frac{\partial q_a}{\partial\varrho}+\frac{\phi P-c-\lambda(P-c)}{P}\Delta\theta\frac{\partial q_a}{\partial\varrho}-(\phi-\lambda)\Delta\theta\frac{\partial q_h}{\partial\varrho} \\ &\quad -(\phi-\lambda)\Delta\theta\left(\frac{\partial q_a}{\partial\varrho}-\frac{\partial q_h}{\partial\varrho}\right)-(P-\Delta\theta(q_a-q_h)) \\ &= -\big(P-\Delta\theta(q_a-q_h)\big)\frac{\partial\lambda}{\partial\varrho}<0,\end{aligned}$$

where we were using the first order conditions (2.48) and (2.49). The final inequality follows, because $\varrho > 0$ and $q_h > 0$ - therefore $P - \Delta\theta(q_a - q_h) = \Delta\theta q_h/\varrho > 0$ - and after applying the implicit function theorem to get

$$\frac{\partial\lambda}{\partial\varrho}=\frac{V''(q_a)V''(q_h)(\Delta\theta(q_a-q_h)-P)}{\frac{(P-c)\varrho}{(1-\phi)P}V''(q_h)(\Delta\theta)^2+\frac{1-\varrho}{1-\phi}V''(q_a)(\Delta\theta)^2}>0. \tag{2.54}$$

Consequently, the solution entails $\varrho^* = 0$ if and only if $\partial\mathcal{V}/\partial\varrho \leq 0$ at $\varrho = 0$.

A necessary condition to get (2.18) is $\phi P > c$. To see this, observe the following

$$
\begin{aligned}
& \frac{(1-\phi)c}{P-c}\big(V(q_l^o) - \theta_l q_l^o - V(\widehat{q}_a) + \theta_l \widehat{q}_a\big) + \phi\Delta\theta(\widehat{q}_a - q_h^{na}) \\
& \quad -(1-\phi)\big(V(\widehat{q}_a) - \theta_h \widehat{q}_a - V(q_h^{na}) + \theta_h q_h^{na}\big) - \frac{\phi P - c}{P-c}P \\
\geq \;& \phi\Delta\theta(\widehat{q}_a - q_h^{na}) - (1-\phi)\big(V(\widehat{q}_a) - \theta_h \widehat{q}_a - V(q_h^{na}) + \theta_h q_h^{na}\big) - \frac{\phi P - c}{P-c}P \\
> \;& \phi\Delta\theta(\widehat{q}_a - q_h^{na}) - (1-\phi)\frac{\phi}{1-\phi}\Delta\theta(\widehat{q}_a - q_h^{na}) - \frac{\phi P - c}{P-c}P = -\frac{\phi P - c}{P-c}P.
\end{aligned}
$$

For the first inequality we used $q_l = q_l^o$. The strict inequality follows from $\widehat{q}_a > q_h^{na}$ and the mean-value-theorem. Hence, for $\phi P \leq c$, the left-hand side of (2.18) is strictly positive. As $P \to \infty$, the left-hand side of (2.18) gets strictly negative, because the first three terms ar bounded, while the last tends to (negative) infinity. this yields existence of P^*. Differentiating the left-hand side of (2.18) with respect to P yields

$$
-\frac{(1-\phi)c}{(P-c)^2}\big(V(q_l^o) - \theta_l q_l^o - V(\widehat{q}_a) + \theta_l \widehat{q}_a\big) - \frac{\phi P - c}{P-c} - P\frac{(1-\phi)c}{P-c} < 0. \tag{2.55}
$$

Consequently, P^* is also unique. □

Proof of Proposition 2.3. The proof of Proposition 2.2 in particular implies, that there exists a unique ϱ^m that maximizes $\mathcal{L}(\varrho)$. Because for given ϱ also the quantities are uniquely defined, we have a unique optimal mechanism.
In the proof of Proposition 2.2 we have already defined the value $\underline{P}^m$ such that (2.16) is binding if and only if $P > \underline{P}^m$ (see equation (2.44) and the following discussion). Hence, the optimal mechanism in this case entails $q_l = q_l^o$ and $q_a = \widehat{q}_a$. □

Proof of Proposition 2.4. If $P > \underline{P}^m$ and $\varrho^m = 1$ we have from (2.16) that $\Delta\theta q_a = P$, hence $q_a = P/\Delta\theta$. It always holds that $q_l = q_l^o$. We next argue that at $\varrho^m > 0$ we have $q_h^o > q_a > q_h \geq q_h^{na}$. To see this, notice that because $q_l = q_l^o$ we have

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \varrho} \leq\; & (\lambda - \phi)\Delta\theta(q_a - q_h) + (1-\phi)(V(q_a) - \theta_h q_a - V(q_h) + \theta_h q_h) \\
& + \frac{\phi P - c}{P-c}P - \lambda P.
\end{aligned} \tag{2.56}
$$

By the mean value theorem we have

$$
(1-\phi)(V(q_a) - \theta_h q_a - V(q_h) + \theta_h q_h) = (1-\phi)(q_a - q_h)(V'(q) - \theta_h)
$$

for some $q \in (q_a, q_h)$. We distinguish the following cases:

- $\lambda \geq 1$. Then $q_h \geq q_a \geq q_l^o$ and because $0 > V'(q_a) - \theta_h \geq V'(q_h) - \theta_h$, we have
$$(1-\phi)(q_a - q_h)(V'(q) - \theta_h) < (\phi - \lambda)\Delta\theta(q_a - q_h). \tag{2.57}$$

- $\lambda \leq \phi$. Here we have $q_h < q_a < q_h^o$ and $V'(q_a) - \theta_h < 0 \leq V'(q_h) - \theta_h$. Consequently (2.57) holds also here.

- $\phi < \lambda < 1$. Then $q_h < q_a$ and $0 > V'(q_h) - \theta_h > V'(q_a) - \theta_h$. Again, this yields (2.57).

Altogether, we have
$$\frac{\partial \mathcal{L}}{\partial \varrho} < P \cdot \left(\frac{\phi P - c}{P - c} - \lambda \right). \tag{2.58}$$

Thus, in a solution with $\varrho^* > 0$ where $\partial\mathcal{L}/\partial\varrho \geq 0$ we must have $\lambda < (\phi P - c)/(P - c) < \phi$. By the first-order conditions for quantities this implies $q_h^{na} \leq q_h < q_a < q_h^o$. Whenever (PC$_l$) is binding, we have $\lambda > 0$ and therefore $q_h > q_h^{na}$. On the fly, this establishes existence of $\overline{P}^m \in [\underline{P}^m, \Delta\theta q_h^o)$ such that $\varrho^m < 1$ for $P > \overline{P}^m$. As long as $\varrho^* = 1$ we have $\Delta\theta q_a \geq P$ by (PC$_l$). Because $q_a < q_h^o$ by what we have shown above, this implies that $\varrho^* < 1$ for $P \geq \Delta\theta q_h^o$. □

Proof of Corollary 2.2. Follows from the arguments given in the main text. □

Chapter 3

Optimal Incentive Contracts to Avert Firm Relocation

This chapter is based on Pollrich and Schmidt (2014).

3.1 Introduction

In a globalized world economy, a firm's location is a strategic choice. Changes in tax regimes, market conditions, or regulations can render production more profitable in one country compared to another, and may induce firms to relocate or outsource production to other countries. Policy makers often perceive such relocation as harmful, because it can cause losses of jobs or reductions in tax revenues. Hence, they sometimes take measures to prevent firms from relocating, or design policies that minimize the incentives for firms to relocate.

We study the issue of firm relocation in a dynamic setting, where a local regulator seeks to prevent the relocation of a firm to some other country in each of two periods. The firm can undertake a location-specific investment that is neither observable to the regulator nor verifiable and, hence, not contractible. The regulator, however, can make transfer payments to the firm contingent on other indicators of the firm's productive activity, such as its output. While the firm's optimal choice of these activities is related to the investment, they are not fully revealing – some activities remain unobservable to the regulator so that the firm's investment cannot be inferred.

We show that a moral hazard problem arises when contracts can only be written on a short-term basis (so that the regulator must offer to the firm a new set of contracts in each period). Such short-term contracting is especially relevant because with changing majorities and legislations, regulators or policy makers may not be able to commit to contractual obligations and future regulations for a sufficiently long period of time. In particular, because firms' location decisions and investments related with them are usually long-term, limited commitment is likely to be a major concern in this context. Limited commitment leads to the problem that the firm can then adopt a 'take-the-money-and-run strategy'. In this case, the firm stays for only one period in its home country and benefits from first-period transfers, but (secretly) lowers its investment, planning to relocate in period 2.[1] We demonstrate that the resulting moral hazard problem leads to distortions in the allocation. In particular, the regulator tightens the regulation in the first period, in order to induce the firm to invest more. Because investment is location-specific, this creates a 'lock-in effect' that averts relocation in both periods under limited commitment.

As a benchmark case, we first consider long-term contracting ('full commitment'). In this case, the regulator can offer contracts to the firm that last for both periods and specify transfers as well as the firm's choice of its (verifiable) production decisions for both periods. The regulator's problem under long-term contracting is simple, because the interests of the regulator and the firm are to some extent aligned. While the firm seeks to maximize its profits, the regulator seeks to avert the firm's relocation at minimal costs, which requires maximal profits. Hence, all productive variables are set to their profit-maximizing levels, and the transfer just compensates the firm for not relocating.

This picture changes drastically under limited commitment, when only short-term contracts can be utilized. On the one hand, the regulator now offers a second-period contract that just compensates the firm for not relocating in that period. On the other hand, the form is now free to relocate in period two. In particular, she can invest little in period one while planning to reject any contract offered in period two. To avoid the latter, a sufficiently high second-period transfer must be promised as a reward for not under-investing. The resulting tension between (1) the need for high second-period transfers to react to the firm's opportunism, and (2) the regulator's parsimony, result-

[1] For example, in 1999 the Finnish telecommunications company Nokia received a subsidy from the German state North Rhine-Westphalia to maintain production of mobile phones in the region. The subsidy was conditioned upon a guarantee to maintain at least 2.856 full-time jobs. Nevertheless, in 2008 Nokia announced plans to shut down production and finally relocated to Romania. For more details see www.spiegel.de/international/germany/the-world-from-berlin-nokia-under-attack-in-germany-a-529218.html.

ing in sequentially optimal second-period transfers, cumulates in an implementation problem.

The first-period contract must be such, that the described tension does not arise in the first place. To achieve this, it must be in the firm's own best interest to stay for both periods when just accepting the first-period contract and, hence, committing to stay for only one period. In this case no second-period contract is required anymore, which solves the problem of sequential-optimality. And because the firm anticipates in period one that she will say also in period two, she will invest accordingly and a 'take-the-money-and-run'-strategy turns out to be not desirable.

The implications of this are twofold. First, the optimal long-term contract is only implementable when relocation is not very attractive. As soon as relocating is sufficiently attractive, the implementation problem gets relevant. This requires a tougher first-period contract to induce already a large investment when the firm just plans to stay for one-period. The lock-in effect of investment then makes it even more attractive for the firm to invest more and stay for both periods. With tougher first-period contracts, the regulator induces the firm to *overinvest*, compared to what would have been optimal under full commitment.

From a policy perspective, our analysis indicates that transfers conditioned only on the location of a firm at a certain point in time (i.e., within a period) may be less effective in averting relocation on a permanent basis than regulations that involve also binding targets for a firm's output or employment. To account for the implementation problem, contracts have to be tougher, in order to induce sufficient investment by the firm. The relocation problem is then solved in period one for all subsequent periods.

To foster intuition, we frame our analysis in the context of the following environmental application.[2] A unilateral introduction of an emission price by a country can induce firms to relocate to other countries with less stringent environmental regulation. Firm relocation is an important channel of 'carbon leakage', or more generally the leakage of emissions to other countries, in response to unilateral emissions regu-

[2] Another application of our model features a regulator, who seeks to induce a pharmaceutical company to develop a new drug. The model developed in this chapter applies if the regulator cannot observe the firm's overall R&D effort to develop the drug, but can subsidize investments in research equipment. Under limited commitment, the firm can pocket any transfers that take place in the first period, and quit the project in the second period.

lation.[3] Even though firm relocation may not be the major cause of leakage,[4] it may be particularly relevant for policy makers given their concern for jobs and international competitiveness. Indeed, the fear of job losses may be one of the reasons why many countries shy away from implementing unilateral climate policies in the first place. In that case, the firm's investment is in abatement capital (or low-carbon technologies) that allows the firm to reduce its operating costs in the light of an emission price in the home country. We refer to the firm's observable activity as 'emissions', while some other activities of the firm (such as output) remain unobservable to the regulator. In this context, transfers can easily be implemented by allocating emission allowances for free during an early phase of a cap-and-trade scheme (Schmidt and Heitzig, 2014). Our results suggest that also in this case, additional criteria (such as firm-specific emission targets) raise the effectiveness of the implicit transfers, unless the regulator is able to commit to future contractual obligations (i.e., transfers) for a sufficiently long period of time.

The remainder of the chapter is organized as follows. Section 3.2 reviews the literature and section 3.3 introduces the model. Section 3.4 studies the relocation problem of the firm in isolation. Regulation is introduced in section 3.5.1, where we analyze the benchmark case of long-term contracting. Section 3.5.2 investigates short-term contracting and contains our main results. Extensions of the model, such as observable but non-contractible investments, and an alternative objective function of the regulator that depends directly on the firm's emissions, are presented in section 3.6. They serve us as a robustness check. Section 3.7 concludes. All formal proofs are relegated to the Appendix.

3.2 Related literature

This chapter contributes to a growing body of literature that tackles the problem of limited commitment in repeated moral hazard problems. E.g., Manso (2011) considers the problem of an agent who is motivated to innovate. The optimal long-term contract that induces the agent to experiment is shown not to be implementable with a sequence of short-term contracts. It is further shown that under certain conditions,

[3] When leakage occurs, the emissions in a foreign country rise in response to the introduction of an emissions control policy in the home country (see, e.g., Babiker (2005)). Leakage is yet another reason why firm relocation may be perceived as harmful by the local regulator.

[4] Another channel is via changes in fossil fuel prices.

outcomes with experimentation completely fail to be implementable. Bergemann and Hege (1998) study the problem of providing venture capital in a dynamic agency model and argue that short-term contracts can never substitute long-term contracts. In their model, however, problems of implementation do not arise. In another paper, Bergemann and Hege (2005) study the funding of a research project with uncertain return and date of completion. Only short-term contracts are considered and a distinction is made between observable and unobservable effort. As opposed to our results, they show that unobservable effort leads to a Pareto-superior outcome, compared to observable investment.

The more general literature on repeated moral hazard is surveyed by Chiappori, Macho, Rey, and Salanié (1994), who derive a principal's optimal contract when motivating an agent to exert costly effort. Rey and Salanié (1990) and Fudenberg, Holmstrom, and Milgrom (1990) provide sufficient conditions for the implementability of the optimal long-term contract via a sequence of short-term contracts. However, they do not characterize the sequence of short-term contracts when the optimal long-term contract is *not* implementable. Fudenberg, Holmstrom, and Milgrom (1990) report two examples for environments where optimal long-term contracts fail to be implementable with short-term contracts, but do not go deeper into this problem.[5]

Our setting also embodies a form of the ratchet effect. Pioneered by Weitzman (1980), the ratchet effect has found its ways into the literature on contracting with limited commitment. Examples are Lazear (1986), Gibbons (1986), Freixas, Guesnerie, and Tirole (1985), and Laffont and Tirole (1988). While Lazear (1986) argues that high-powered incentives can overcome the ratchet effect, Laffont and Tirole (1988) prove a result on the impossibility of implementing full separation with a continuum of types. All these works study models of adverse selection. The issue is then to compensate the agent today for being exploited in the future, because ex-ante private information is typically revealed over time. We instead study a model of moral hazard, where the exploitation in the future has severe consequences on the problem of implementing effort in the first place.

A recent paper that studies the ratchet effect in a model with moral hazard is Bhaskar (2014). He studies a dynamic principal-agent problem with moral hazard and learning. The difficulty of the job, undertaken by the agent, is a priori unknown

[5]Our model can be seen as a version of Example 1 in Fudenberg, Holmstrom, and Milgrom (1990). The intuition behind their Example 2, however, fits better with the observed implementation problem in this chapter.

to both parties. Conditional on first-period effort and output, both principal and agent update their beliefs. When shirking, the agent's posterior differs from the principal's, which gives rise to a ratchet effect that leads to a failure of implementability that is similar to the one presented in this chapter. The agent can adopt a 'take-the-money-and-run' strategy, which makes deviations from interior values profitable.

Due to the commitment problem under short-term contracting, the chapter is also related to the literature on incomplete contracts, e.g. Hart and Moore (1988). As in this literature, we allow contracts to depend on some observable characteristics, but not on investments. Our analysis of short-term contracting establishes a new channel for a contractual hold-up: although the contracts we analyze are rich enough to mitigate hold-up within a period (or under full commitment), the threat of exploitation in future periods resurrects the hold-up problem under limited commitment. As compared to the classical results in that literature (see Che and Sákovics (2004) for an overview), we identify over-investment as another possible consequence of incomplete contracting. Joskow (1987) finds empirical evidence for a link between the contractual commitments of future trade and importance of relationship-specific investment. The chapter provides a theoretical foundation: when the contract length falls short of the time in which investments are recouped, efficient investment cannot be implemented.

In a model of repeated climate contracting between countries, Harstad (2012) finds results that are related to ours. Countries repeatedly negotiate climate contracts that specify emission levels. Between the contracting stages they invest in abatement technology. The author finds that shorter contract duration leads to tougher contracts and lower emission levels are agreed upon. However, investments remain at an inefficiently low level, whereas in our model contracts are tougher *and* investments are inefficiently high.

The problem of firm relocation has been studied in different strands of literature. Horstmann and Markusen (1992), e.g., study the impact of a trade policy on market structure. They report that 'small policy changes can produce large welfare effects when equilibrium market structure shifts'. Also tax competition in general affects firm location, Wilson and Wildasin (2004) and Bucovetsky (2005) provide an overview.[6] The impact of unilateral environmental regulation on firms' location decisions was first analyzed formally by Markusen, Morey, and Olewiler (1993).[7] In a two-country

[6] See also Haufler and Wooton (2010).

[7] See also Markusen, Morey, and Olewiler (1995). Other examples include Motta and Thisse (1994), who analyze the relocation of firms already established in their home country in response to a unilateral

model, firms decide where to locate after governments have determined environmental taxes. Firms' location decisions are, therefore, very sensitive to differences in tax policies, as confirmed by Ulph (1994) in a numerical calibration of the model. The chapter complements this literature in that it provides a method to counterbalance the adverse effects on firm location.

Schmidt and Heitzig (2014) study the dynamics of 'grandfathering' schemes. They show that such transfer schemes can permanently avert firm relocation even when they terminate in finite time. In contrast to what we do in this chapter, full contractual commitment by the regulator is assumed. Their findings conform with our results on long-term contracting. In particular, with full commitment, simple transfer schemes are sufficient, as the regulator need not interfere directly with the firm's productive decisions. The promise of transfers that last for a sufficiently long period of time induces the optimal investment by the firm, and permanently averts its relocation. With limited commitment, however, our results indicate that these simple grandfathering schemes are no longer optimal and contracts should be made contingent on other observable characteristics, such as emissions.

3.3 Model

3.3.1 The firm

We analyze the following two-period model: There is one firm that is initially located in country A, where it earns per-period profits of $\pi_A(e, a)$. The variable e reflects some productive activity, and a is the stock of capital available to the firm. For illustrative purposes, we will interpret these variables in terms of our environmental example (motivated in the introduction) throughout the chapter. Then e stands for the firm's *emissions*, and a is the firm's stock of *abatement capital*. Note that the profit function $\pi_A(e, a)$ is given in a *reduced form*. In particular, all other potential factors (e.g. input and output quantities, prices) are always chosen optimally by the firm, for any given values of e and a. Below, we show how to derive the firm's profit in the reduced form $\pi_A(e, a)$ in a specific example.

Emission levels are chosen by the firm in each period, and we denote e_τ the emis-

anti-pollution policy pursued by the government in their home country. Further, Ulph and Valentini (1997) analyze strategic environmental policy in a setting where different sectors are linked via an input-output relation.

sion level in period $\tau \in \{1, 2\}$. The capital stock a is established at the beginning of period 1 and is thereafter available for both periods of production.[8] We further assume that abatement capital is immobile, i.e. it can only be utilized in country A.[9] The cost of installing a capital stock of $a \geq 0$ is given by the strictly convex cost function $K(a)$, with $K(0) = K'(0) = 0$. The firm's discounted profit from producing in country A in *both* periods, when choosing emission levels e_1 and e_2 as well as capital a is, therefore,

$$\pi_A(e_1, a) - K(a) + \delta \pi_A(e_2, a), \tag{3.1}$$

where $\delta > 0$ is the discount factor.[10]

We assume that at the beginning of each period, the firm has the possibility to relocate to some other country, in the following referred to as 'country B'. In country B, the firm earns a fixed per-period profit of π_B.[11] Relocation is once and for all, and for simplicity assumed to be costless. If the firm relocates immediately (i.e. at the beginning of period 1) to country B, it earns a total profit of

$$V_B = (1 + \delta)\pi_B. \tag{3.2}$$

In this case, the firm has no incentive to invest in abatement capital. The firm can also stay in A for only one period, and relocate to B at the beginning of period 2. This strategy, referred to as 'location plan AB', amounts to a discounted profit of

$$\pi_A(e_1, a) - K(a) + \delta \pi_B. \tag{3.3}$$

We use the following technical assumptions regarding the profit function $\pi_A(e, a)$, defined on an open interval $(\underline{e}, \overline{e})$, with $-\infty \leq \underline{e} < \overline{e} \leq \infty$.[12]

Assumptions:

(A1) $\pi_A(e, a)$ is strictly concave in e, and for all $a \geq 0$ we have $\frac{\partial \pi_A}{\partial e} = +\infty$ for

[8]In particular, we assume away depreciation. Allowing for a positive rate of depreciation would, however, not change our main results.

[9]Examples include investments in more energy-efficient production technologies, or investments in physical capital such as a building.

[10]We allow for $\delta > 1$, which admits time periods of different length and/or economic importance.

[11]In the context of our environmental example, country B may, e.g., be a country that does not regulate emissions. Hence, even if capital were mobile, a prior investment in abatement capital does not affect the firm's profit after relocation.

[12]Negative values for e can be interpreted as selling emission rights on the market.

$e \to \underline{e}$, as well as $\frac{\partial \pi_A}{\partial e} < 0$ for $e \to \overline{e}$.

(A2) $\pi_A(e,a)$ is strictly concave in a, and $\frac{\partial \pi_A}{\partial a} > 0$ holds at $a = 0$ for all $e \in (\underline{e}, \overline{e})$; furthermore, $\frac{\partial \pi_A}{\partial a}$ is bounded from above for all $e \in (\underline{e}, \overline{e})$.

(A3) The Hessian of $\pi_A(e,a)$ is negative definite.

(A4) $\frac{\partial^2 \pi_A}{\partial e \partial a} < 0$.

(A5) $\exists\, \varepsilon > 0$ such that whenever $\partial \pi_A/\partial e = 0$ then $\partial \pi_A/\partial a > \varepsilon$.

The first three assumptions are technical: (A1) states that $\pi_A(e,a)$ is a regular profit function in e for all possible values of a (i.e., there exists a unique interior maximizer) and rules out boundary solutions for e. Assumption (A2) implies that investment exhibits diminishing returns, and it is never optimal to choose $a = 0$ (unless the firm relocates immediately). (A3) guarantees concavity of implicitly defined functions (such functions are introduced later on).

The last two assumptions describe the relation between emissions and investment: (A4) is a single-crossing property, and implies that emissions and investment are substitutes.[13] Assumption (A5) implies that whenever the firm is free to choose e optimally, it is always better off with a larger capital stock when the investment costs in a are ignored.

Example 3.1. *Consider a polluting firm that produces an output quantity q, emitting e units of greenhouse gases. The firm faces the inverse demand $P(q) = 3 - q/2$. Marginal costs of production are constant and normalized to zero. The emissions price in A (e.g., following the introduction of a cap-and-trade scheme) is equal to 1 in both periods. Consequently, the firm's per-period profit in country A, gross of abatement capital installation cost, is*

$$\tilde{\pi}_A(e,q) = (3 - q/2)q - e.$$

Emissions are a function of output and the firm's abatement capital stock. For simplicity, we assume that the firm's emissions are additive in q and a, i.e. $e(q,a) = q - a$. Inserting this into $\tilde{\pi}_A(e,q)$, we obtain the firm's profit function in the reduced form:[14]

$$\pi_A(e,a) = 3a + 2e - (a+e)^2/2. \tag{3.4}$$

[13] Intuitively, if the firm has a larger abatement capital stock then its *optimal* emissions are lower.

[14] It is easy to verify that the function $\pi_A(e,a)$ fulfills our earlier assumptions.

We will return to this simple example frequently throughout the chapter, in order to illustrate our findings.

3.3.2 The regulator

In country A a regulator (or policy maker) is concerned with the firm's option to relocate. In particular, as soon as the firm relocates, welfare in country A is reduced by some fixed amount $L > 0$, e.g., due to job losses or lower tax revenues.[15]

Because of the potential loss L, the regulator's main interest is to avert relocation of the firm on a permanent basis. To this end, the regulator offers to the firm contracts in a take-it-or-leave-it manner. We assume that the firm's emissions in each period are contractible. However, the investment in abatement capital is neither observable to the regulator nor verifiable. Contracts thus specify a location-specific transfer to the firm, denoted by t, and emission levels that the firm has to comply with (in order to obtain the transfer). The firm can reject any contract offer and either relocate to country B, or produce in country A at its own, un-subsidized expense.

The regulator maximizes the following welfare function

$$W = -\chi_1 t_1 - \chi_2 \delta t_2 - (1 - \chi_2) L, \qquad (3.5)$$

where $\chi_\tau = 1$ if the firm operates in country A in period τ (and accepts the contract offered in that period), and $\chi_\tau = 0$ otherwise.[16] The regulator and the firm use the same discount factor $\delta > 0$.

Throughout this chapter, we distinguish between long-term and short-term contracts. The former specify emission levels and transfers for each individual period, i.e. a long-term contract is a quadruple (t_1, e_1, t_2, e_2). This implicitly assumes that the regulator can fully commit to all present and future contractual obligations. Commitment here is two-sided, i.e. also the firm, after signing the contract, is committed to

[15] The assumption that L is independent of whether the firm relocates in period 1 or in period 2 highlights the regulator's interest in averting relocation on a *permanent* basis (rather than on a temporary one). In our environmental example, an emission price is implemented by some higher authority (e.g., on the federal level), while transfers are paid by a local regulator who's primary objective it is to avert the firm's relocation. Hence, the firm's emissions do not directly affect the regulator's payoff. In Section 3.6, we introduce an alternative payoff function for the regulator that also depends on the firm's choice of e, as well as on the period in which the firm relocates. The regulator may then also benefit from averting relocation only in period 1. We will show that our main results are unaffected by these changes.

[16] Because relocation is by assumption irreversible, $\chi_2 = 1$ requires that also $\chi_1 = 1$ holds.

staying in country A throughout the contract duration.[17]

The timing with long-term contracting is as follows. First, the regulator offers a contract. After observing the contract offer, the firm decides whether or not to relocate to country B. If the firm relocates, the game ends. Otherwise, it decides whether or not to accept the contract, and chooses a level of abatement capital investment.[18] Finally, production starts under the terms specified in the contract or, in case no contract was signed, the firm chooses its productive variables.

Under short-term contracting neither the regulator nor the firm have the ability to make commitments that last for more than one period.[19] Hence, the regulator resorts to a sequence of spot contracts (t_τ, e_τ). The timing for this case is as follows.

1. Regulator offers contract (t_1, e_1).
2. Firm accepts/rejects and location choice A/B.
3. Firm chooses a and produces e_1. Transfer t_1 paid to the firm.
4. Regulator offers contract (t_2, e_2).
5. Firm accepts/rejects and location A/B.
6. Firm produces e_2. Transfer t_2 paid to the firm.

In the first period a short-term contract (t_1, e_1) is offered to the firm. After observing the contract, the firm decides on its location and whether or not to accept the contract. The game ends whenever the firm relocates. Otherwise, the firm invests in abatement capital and production takes place (according to the terms specified in the contract if accepted). At the end of period 1, the transfer is paid to the firm, in case it accepted the contract. Period 2 starts with a new contract offer (t_2, e_2) by the regulator (unless the firm already relocated in period 1). The firm observes the offered contract and decides whether or not to relocate in period 2. If it stays in A, the firm can accept the contract and produce according to the contractual terms or reject the contract, in which case it produces on its own account and does not receive any transfer payment in period 2. Again, the transfer is paid at the end of period 2.

[17] This formulation rules out contracts that keep the firm for the first period and impose relocation in the second period. Because such contracts are never desirable, their exclusion is without loss.

[18] The firm's decisions within a period are, of course, simultaneous.

[19] Intermediate cases of one-sided commitment are simple in our model. E.g., if the regulator has full commitment power but not the firm, postponing all transfers to period 2 – after the option to relocate has vanished – is sufficient to implement the full commitment outcome.

3.3.3 Equilibrium concept

We argue in the following that even though we study a dynamic game with imperfect information, we can use *Subgame Perfect Nash Equilibrium* (SPNE) as our solution concept, and hence, backward induction as solution method. This is obvious in the case of long-term contracting where the regulator moves only once and all remaining decisions are taken by the firm (after observing the long-term contract offered by the regulator).

Under short-term contracting, there is no proper subgame after stage 3 (see above), because the regulator does not observe the firm's choice of a. However, stages 5 and 6 constitute a proper subgame, because the firm has perfect recall. Furthermore, the sequentiality of stages 3 and 4 (firm's choice of a and second-period contract offer (t_2, e_2)) is inconsequential for the equilibrium outcome because no information is revealed between these two stages. Hence, we can effectively treat these two stages as *simultaneous* moves. This allows us to solve the game by backward induction.[20]

Furthermore, throughout the main part of the chapter we focus on pure strategies. This is clearly without loss of generality when we analyze long-term contracts. With short-term contracting, randomization could be beneficial when the firm chooses its investment. However, as we formally prove in Appendix 3.B, there are no additional equilibria in mixed strategies. Hence, focusing on pure strategy equilibria is without loss of generality also in the case of short-term contracting.

3.4 Preliminaries and the 'no-regulation' benchmark

In this section we consider the firm's problem in isolation and identify conditions under which relocation occurs. It will turn out convenient to use the following shorthand notations. Let

$$\pi_A^*(a) = \max_e \pi_A(e, a), \tag{3.6}$$

be the firm's *maximal profit* in one period after having installed capital stock a. Denote $e^*(a)$ the corresponding level of emissions. Using this, we can define

$$V_A(e_1) = \max_a \big(\pi_A(e_1, a) - K(a) + \delta\pi_A^*(a)\big). \tag{3.7}$$

[20]The alternative is to use Perfect Bayesian Nash Equilibrium. This requires specifying beliefs of the regulator in stage 4 about the firm's choice of investment. Because of the simple structure, these PBNE correspond to the SPNE.

This represents the firm's discounted profit when staying in country A in both periods, with first-period emissions *fixed* (e.g., in a contract) at a level of e_1, while choosing e_2 optimally in period 2, and choosing a optimally in period 1. The corresponding optimal level of investment is denoted by $a_A(e_1)$. Similarly,

$$V_{AB}(e_1) = \max_a \big(\pi_A(e_1, a) - K(a) + \delta\pi_B\big), \tag{3.8}$$

is the firm's profit under location plan AB with first-period emissions e_1, given an optimal investment for this location plan. The corresponding maximizer is denoted by $a_{AB}(e_1)$. The following Lemma states properties of these functions and their maximizers.

Lemma 3.1. *(1)* $e^*(a)$ *is unique and strictly decreasing,*

(2) $\pi_A^*(a)$ *is strictly increasing, concave, and* $\lim_{a\to\infty} \pi_A^*(a) = +\infty$*,*

(3) $a_A(e_1)$ *and* $a_{AB}(e_1)$ *are unique and strictly decreasing,*

(4) $V_A(e_1)$ *and* $V_{AB}(e_1)$ *are strictly concave and have unique maximizers,*

(5) $a_A(e_1) > a_{AB}(e_1)$ *for all* $e_1 \in (\underline{e}, \overline{e})$*.*

The first result confirms that a firm that has installed a larger abatement capital stock optimally chooses lower emissions. The second result rephrases our earlier assumption (A5) that $\partial\pi_A/\partial e = 0$ implies $\partial\pi_A/\partial a > 0$ and provides a first indication towards a lock-in effect, namely a sufficiently large investment renders relocation unprofitable even for large values of π_B. The functions $a_A(e_1)$ and $a_{AB}(e_1)$ are decreasing because in our model a stricter regulation in the first period corresponds to a *smaller* value of e_1 (emissions are regulated more tightly). Accordingly, the firm responds with a larger investment when e_1 is smaller (both under location plan AB or when the firm plans to stay permanently in A). The final result says that if the firm plans to stay in A in both periods, it invests more than when it plans to relocate after one period.

The next lemma is an immediate consequence of the investment cost being sunk.

Lemma 3.2. *For any level of first-period emissions, the option to relocate after one period is always inferior to either immediate relocation or no relocation (or both). More specifically, it holds for any* e_1 *that* $V_{AB}(e_1) < \max\{V_A(e_1), V_B\}$*.*

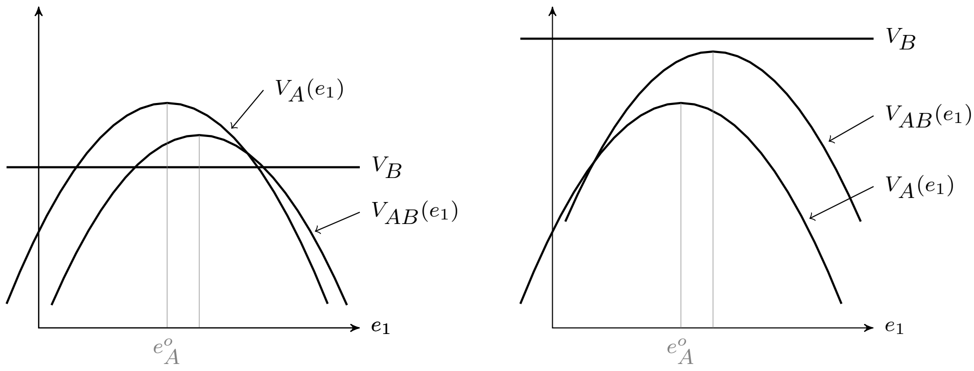

Figure 3.1: Profit functions $V_A(e_1)$, $V_{AB}(e_1)$, and V_B for low π_B (left) and high π_B (right).

The Lemma establishes a *lock-in* effect. Whenever the firm finds it optimal to stay for one period in country A, it will undertake some investment. Intuitively, location plan AB can only be optimal if the net profit in period 1, i.e. profit from production minus the cost of installing capital, exceeds the profit in country B. But then the corresponding per-period profit of production in country A, gross of investment costs, clearly exceeds π_B when the firm implements $a_{AB}(e_1)$. So it must be profitable for the firm to stay in country A also for the second period. By raising its investment to the level $a_A(e_1)$, the firm can achieve an even higher profit.

Figure 3.1 illustrates the typical shape of the firm's profit function for the different location plans. Note, that raising π_B does not affect V_A, whereas it shifts V_{AB} as well as V_B upwards.

According to Lemma 3.2, the firm prefers either to stay in country A for both periods or to relocate immediately. Only the latter case is of interest for us, since it calls for regulatory intervention. To make this more precise, let e^o_A be the optimal (first-period) emission level when the firm plans to stay in country A for both periods. It is given by

$$e^o_A = arg \max_{e_1} V_A(e_1). \tag{3.9}$$

Because the firm uses the same capital stock in each period, it is straightforward to verify that given this optimal choice of first-period emissions, it holds that $e_2 = e_1 = e^o_A$ if the firm is free to choose its emissions in period 2. Define $V^o_A := V_A(e^o_A)$ and $a^o_A := a_A(e^o_A)$. The following lemma is an immediate consequence of the preceding derivations.

Lemma 3.3. *Absent regulatory intervention, the firm strictly prefers immediate relocation whenever $\pi_B > \pi_B^o$, and no relocation otherwise. The critical value π_B^o is given by*

$$\pi_B^o := \frac{V_A^o}{1+\delta}. \tag{3.10}$$

Throughout the rest of the chapter we maintain the assumption $\pi_B \geq \pi_B^o$. Hence, in the absence of regulatory intervention the firm relocates immediately.

Example 3.2. *Maximizing $\pi_A(e,a) = 3a + 2e - (a+e)^2/2$ over e, we find that the firm's optimal emissions (given a) are $e^*(a) = 2 - a$. Therefore $\pi_A^*(a) = 2 + a$. Let investment costs be given by the quadratic cost function $K(a) = a^2/2$. If the firm plans to stay in country A in both periods, and is constrained to emit (no more than) e_1 units in period 1 (e.g., by the regulator), it thus solves:*

$$\max_a \; 3a + 2e_1 - \frac{(a+e_1)^2}{2} - \frac{a^2}{2} + \delta(2+a).$$

This yields $a_A(e_1) = (3 - e_1 + \delta)/2$ and $V_A(e_1) = \frac{1}{2}(5+\delta)(1+\delta) - \frac{1}{4}(e_1 - (1-\delta))^2$. The latter implies $e_A^o = 1 - \delta$ and $V_A^o = \frac{1}{2}(5+\delta)(1+\delta)$. The critical level of π_B for relocation is $\pi_B^o = \frac{1}{2}(5+\delta)$. If the firm plans to stay in country A for only one period, it solves:

$$\max_a \; 3a + 2e_1 - \frac{(a+e_1)^2}{2} - \frac{a^2}{2} + \delta\pi_B.$$

This yields $a_{AB}(e_1) = (3 - e_1)/2$, and $V_{AB}(e_1) = \frac{5}{2} - \frac{1}{4}(e_1 - 1)^2 + \delta\pi_B$. The firm's optimal choice of first-period emissions is $e_{AB} = 1$. Observe that the firm's emissions are higher and the abatement capital investment is smaller when it plans to relocate after one period (we find $a_A^o = 1 + \delta$ and $a_{AB} = 1$).

3.5 Regulation

This section studies the optimal regulatory policy in the presence of the threat of firm relocation. We first analyze the benchmark case of long-term contracting (full commitment), and then proceed to short-term contracting (limited commitment).

3.5.1 Long-term contracting

The regulator's payoff from not offering a contract is $-L$. Alternatively, the regulator can offer a long-term contract that requires the firm to produce in country A in both periods. In finding the optimal contract that permanently averts relocation, the regulator solves the following program

$$
\begin{aligned}
&\min_{t_1,e_1,t_2,e_2,a} \quad t_1 + \delta t_2 \\
\text{s.t.} \quad & t_1 + \pi_A(e_1,a) - K(a) + \delta\big(t_2 + \pi_A(e_2,a)\big) \geq V_B \text{ , and} && \text{(PC)} \\
& t_1 + \pi_A(e_1,a) - K(a) + \delta\big(t_2 + \pi_A(e_2,a)\big) \geq \\
& \qquad t_1 + \pi_A(e_1,\widetilde{a}) - K(\widetilde{a}) + \delta\big(t_2 + \pi_A(e_2,\widetilde{a})\big) \ \ \forall \widetilde{a}. && \text{(MH-1)}
\end{aligned}
$$

The participation constraint (PC) ensures that the firm prefers accepting the contract (and not relocating) to immediate relocation. Constraint (MH-1) is a moral hazard constraint, that ensures the firm chooses the *intended* level of investment. Because we assume two-sided commitment, the distribution of transfers across periods is inconsequential and we can substitute for the total transfer $t = t_1 + \delta t_2$.[21]

Obviously the participation constraint (PC) is binding. Together with the moral hazard constraint (MH-1) the minimal (total) transfer t that is required to avert relocation in both periods when emissions are chosen at levels e_1 and e_2 is

$$t = V_B - \max_a \big(\pi_A(e_1,a) - K(a) + \delta\pi_A(e_2,a)\big). \tag{3.11}$$

The regulator's minimization program given above, therefore, corresponds to minimizing (3.11) with respect to e_1 and e_2. This is equivalent to maximizing $V_A(e_1)$ over e_1, which yields $e_1 = e_A^o$ as defined in (3.9). The minimal total transfer required to avert relocation is, therefore, $t = V_B - V_A^o$, and the regulator, accounting for the welfare loss from relocation, offers a contract that averts relocation if and only if this transfer does not exceed L. The following proposition summarizes.

Proposition 3.1. *The optimal long-term contract specifies $e_1 = e_2 = e_A^o$, pays a total transfer of $t^o := V_B - V_A^o$ and the firm does not relocate, whenever $L \geq t^o$. Otherwise, the regulator offers the* null contract *and the firm relocates immediately.*

Notice the following alternative way of implementing the optimal long-term contract: Because of full commitment on the side of the firm, the regulator can simply offer the lump-sum subsidy t^o for the firm's commitment not to relocate in any of the two periods. This leaves the optimal choices of e_1 and e_2 at the firm's discretion. The

[21]This also relies on the assumption that regulator and firm have a common discount factor. With differing discount factors, the regulator would have a preference for either paying all transfers in period 1 or in period 2.

firm then chooses emissions and investment so as to maximize its discounted profit from two periods of production in country A. But this implies $e_1 = e_2 = e_A^o$ and $a = a_A^o$, as we have shown in Section 3.4. Acceptance of the subsidy t^o is implied by its definition in Proposition 3.1. Hence, under full commitment, a simple location-based subsidization is sufficient to avert firm relocation with minimal transfers; the regulator does not need to interfere directly with the firm's productive activities.

Example 3.3. *Applying Proposition 3.1, the optimal long-term contract specifies emission targets $e_1 = e_2 = e_A^o = 1 - \delta$. The firm's discounted profit in A is $V_A^o = \frac{1}{2}(1+\delta)(5+\delta)$, and a total transfer of $t^o = V_B - V_A^o = (1+\delta)\big[\pi_B - \frac{1}{2}(5+\delta)\big]$ is required to avert relocation. From the expression for t^o we also get $\pi_B^o = \frac{1}{2}(5+\delta)$.*

3.5.2 Short-term contracting

We move on to the study of short-term contracting. Hence, we assume that the regulator cannot commit to a contract that specifies emissions and transfers for both periods and instead resorts to a sequence of short-term contracts. Also the firm cannot commit in period 1 to not relocating in period 2.

For each contracting party, limited commitment generates a new constraint. First, the regulator's second-period contract offer must be sequentially optimal. In particular, the second-period transfer can be no higher than required to avert relocation in that period. Second, the firm has the option to accept the first-period contract but nevertheless relocate in period 2. In order to prevent this, a sufficiently large second-period transfer has to be 'promised'. We show in this section that these two restrictions can only be compatible if the latter constraint is irrelevant, i.e. upon accepting the first-period contract the firm already prefers to stay for two periods and planned relocation is inferior. To achieve this, the regulator sets a stringent (i.e. low) first-period emission target e_1. This induces a lock-in that prevents relocation in both periods without transfers in period 2.

As before, offering no contract results in a welfare of $-L$. The relevant alternative to the null contract in period 1 is a contract offer that is accepted by the firm, and leads to an outcome where the firm does not relocate in period 2. Acceptance of the first-period contract, while taking the continuation play as given, is induced by constraint (PC). Similarly, the constraint for the firm choosing investment a provided it accepts the respective contracts in each of the two periods is given by (MH-1).

Furthermore, after having installed capital stock a in period 1, the firm is willing to accept the second-period contract offer (t_2, e_2) if and only if $t_2 + \pi_A(e_2, a) \geq \max\{\pi_B, \pi_A^*(a)\}$. Note that the firm has the option to produce in A at its own expense, earning a maximal profit of $\pi_A^*(a)$, which leads to zero transfers for large values of a. Hence, the second-period contract (t_2, e_2) is sequentially optimally provided the firm invests a, whenever

$$t_2 = \max\{0, \pi_B - \pi_A^*(a)\}, \qquad e_2 = e^*(a). \tag{SO}$$

As in the case of long-term contracting, the regulator's and the firm's interests are to some extent aligned: minimizing the transfer payment, the regulator seeks to maximize the firm's profit over e_2. What is crucial is that whenever $t_2 > 0$, this transfer just compensates the firm for not relocating in period 2. However, if $\pi_A^*(a) \geq \pi_B$, then no second-period transfer is required.[22]

The other new constraint concerns the firm's possibility to (secretely) plan relocation. Doing so, after having accepted the first-period contract (t_1, e_1), the firm chooses investment $a_{AB}(e_1)$, and earns a discounted profit of $t_1 + V_{AB}(e_1)$. This leads to the additional moral hazard constraint

$$t_1 + \pi_A(e_1, a) - K(a) + \delta\big(t_2 + \pi_A(e_2, a)\big) \geq t_1 + V_{AB}(e_1). \tag{MH-2}$$

The regulator's problem of finding the minimal transfer(s) that permanently avert relocation can, therefore, be stated as follows:

$$\min_{t_1, e_1, t_2, e_2, a} \; t_1 + \delta t_2, \qquad \text{subject to (PC), (MH-1), (MH-2), (SO).} \tag{$\mathcal{P}_S$}$$

Before solving problem $\mathcal{P}_S$, let us first characterize the set of first-period emission levels that induce an equilibrium in the continuation game where the firm never relocates. Hence, we are looking for levels of e_1 for which there *exists* a contract (t_2, e_2) and an investment level a such that constraints (MH-1),(MH-2), and (SO) are satisfied. Notice that constraint (SO) essentially pins down (t_2, e_2) for a *given* level of investment a. Similarly, for a *given* second-period emission level e_2, we can derive a from constraint (MH-1).[23] Using the latter condition and $e_2 = e^*(a)$, we can thus

[22] We assume that when $\pi_A^*(a) \geq \pi_B$, the firm still accepts a contract offer with $t_2 = 0$ and emissions at the level $e_2 = e^*(a)$.

[23] This is also what we have done when deriving the optimal long-term contract. Note, that any

rewrite constraint (MH-2) as follows

$$\delta t_2 + V_A(e_1) \geq V_{AB}(e_1). \tag{MH-2$'$}$$

In this representation the role of the second-period transfer becomes clear. When investing, the firm faces two options: Either it invests little and relocates in period 2, rejecting the second-period contract offer. Or it invests more, planning to stay in A in both periods and accepting the second-period contract offer. Because the actual investment level is not observable to the regulator, the second-period offer cannot be made contingent on it. When seeking to implement an outcome where the firm never relocates, the second-period contract offer (t_2, e_2) is implicitly contingent on the optimal investment level for the second option (no planned relocation), by conditions (MH-1) and (SO). But the resulting second-period transfer has to compensate the firm also for not secretly under-investing, i.e., by condition (MH-2′), it has to hold that $\delta t_2 \geq V_{AB}(e_1) - V_A(e_1)$. The following result shows that this condition restricts the range of implementable outcomes.

Proposition 3.2. *For a first-period emission level e_1, there exists a second-period contract (t_2, e_2) and an investment level a such that constraints* (MH-1)*,* (MH-2)*, and* (SO) *are satisfied if and only if $V_A(e_1) \geq V_{AB}(e_1)$.*

If the condition in the proposition is met, constraint (MH-2) has no bite. This can be seen best from its reformulation into (MH-2′). Provided that $V_A(e_1) \geq V_{AB}(e_1)$, any non-negative transfer t_2 satisfies the constraint. If, however, $V_A(e_1) < V_{AB}(e_1)$, constraint (MH-2′) imposes a lower bound on t_2, as argued above. Intuitively, in order to satisfy constraint (MH-2′), the second-period transfer not only has to account for the difference in second-period profits, but also for the respective difference in first-period profits that arises when the firm plans to stay in A in both periods, rather than to relocate after period 1. In particular, because the underlying investments differ in the two cases, first-period profits are strictly higher with planned relocation compared to no relocation, and the second-period transfer – serving as reward – has to compensate for this difference. However, because the regulator has no commitment power, offering such a reward is not credible. Any sequentially optimal second-period transfer, i.e., any t_2 that satisfies (SO) only compensates the firm for not relocating within that period, and fails to take into account investment costs that were incurred prior to this

combination e_1, e_2 leads to a unique investment level.

period.

Notice a crucial consequence of Proposition 3.2: the condition $V_A(e_1) \geq V_{AB}(e_1)$ implies that no second-period transfer is required to avert relocation in period 2. In other words, an equilibrium with no relocation under short-term contracting necessarily implies a situation where the firm is *locked-in* after the first period.

Proposition 3.2 also allows us to determine when the optimal *long-term* contract is implementable via a sequence of short-term contracts:

Corollary 3.1. *The optimal long-term contract can be implemented via a sequence of short-term contracts if and only if* $V_A^o \geq V_{AB}(e_A^o)$*. This is equivalent to* $\pi_B \leq \pi_B^\sharp$*, where*

$$\pi_B^\sharp := \tfrac{1}{\delta}\big(V_A^o - \pi_A(e_A^o, a_{AB}(e_A^o)) + K(a_{AB}(e_A^o))\big) > \pi_B^o.$$

The respective sequence of contracts entails $(t_1, e_1) = (t^o, e_A^o)$*, and* $(t_2, e_2) = (0, e_A^o)$*.*

We now proceed with the analysis of optimal short-term contracts when $\pi_B > \pi_B^\sharp$. The following result makes the analysis more transparent, by mapping the condition $V_A(e_1) \geq V_{AB}(e_1)$ from Proposition 3.2 to a line segment.

Lemma 3.4. *Assume* $\pi_B > \pi_B^\sharp$*. Then there exists a unique value* $e^\sharp$*, with* $\underline{e} < e^\sharp < e_A^o$*, such that* $V_A(e_1) \geq V_{AB}(e_1)$ *holds if and only if* $e_1 \leq e^\sharp$*. The level* $e^\sharp$ *decreases with* π_B*.*

Hence, only sufficiently low emission targets for the first period can be utilized to implement an outcome without relocation in any period. By offering more high-powered incentives in the first period, the regulator enforces a sufficiently high abatement capital investment by the firm. This renders the relocation option in period 2 unprofitable when the firm optimally exploits its possibilities to invest in abatement capital. Planning to relocate after period 1 is, then, no longer optimal from the firm's perspective, because staying for only one period in A already involves a fairly large investment. The firm then prefers to invest even more, and realizes the rents from the investment also in period 2.

Finding the optimal first-period contract, i.e. the first-period emission level e_1 that implements an equilibrium where the firm stays for both periods in country A with the lowest (total) transfers, is now straightforward. Because $V_A(e_1)$ is strictly concave, implementing $e_1 = e^\sharp$ leads to lowest transfers and is, therefore, optimal. Regarding the cost of implementing such an outcome, the total transfer required is given by

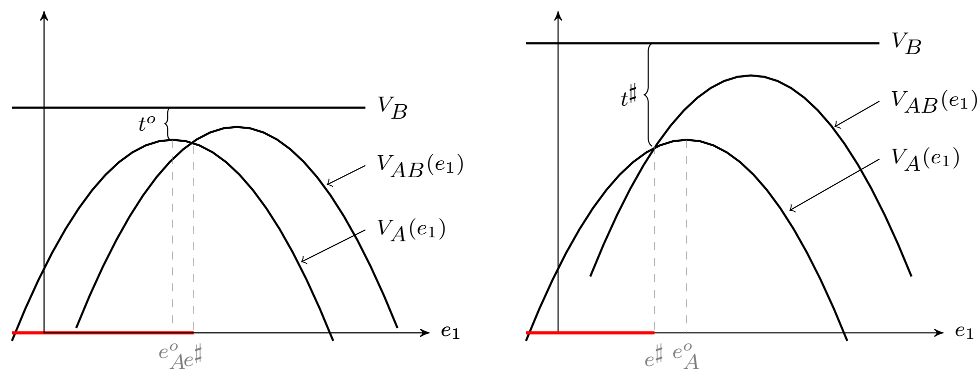

Figure 3.1: Optimal first-period contracts with short-term contracting; left: $e^o_A < e^\sharp$, right: $e^o_A > e^\sharp$. Implementable levels of e_1 are shown in red.

$t_1 = V_B - V_A(e^\sharp)$, and the regulator prefers this to immediate relocation whenever $t_1 \leq L$.

Proposition 3.3. *With short-term contracting the optimal first-period contract is*

- $(t_1, e_1) = (t^o, e^o_A)$, *if* $\pi_B \leq \pi^\sharp_B$ *and* $L \geq t^o$*;*
- $(t_1, e_1) = (t^\sharp, e^\sharp)$, *if* $\pi_B > \pi^\sharp_B$ *and* $L \geq t^\sharp$, *with* $t^\sharp := V_B - V_A(e^\sharp) > t^o$*;*
- *the null contract otherwise.*

In the first two cases the second-period contract is $(t_2, e_2) = (0, e^*(a_A(e_1)))$.

The implications of Proposition 3.3 are as follows: For moderate relocation profits π_B, the lack of commitment has no consequence for the optimal contract. Both with long-term and with short-term contracting, a transfer has to paid only in period 1, and the firm invests enough so that relocation in period 2 is no longer in its interest. Hence, for moderate values of π_B a one-period contract is sufficient to resolve the relocation problem on a permanent basis, even without regulation in period 2. This case is depicted in the left panel of Figure 3.1. Observe that at $e_1 = e^o_A$, it holds that $V_{AB}(e_1) < V_A(e_1)$. Hence, as the firm has to comply with the emission target e_1 in order to obtain the transfer t_1 in the first period, the option to relocate in period 2 is effectively ruled out.

However, when the outside option in form of the relocation option is more attractive, limited commitment affects the design of the optimal contract in period 1, and the effect can be severe. A tension arises between the regulator's parsimony, i.e., offering a sequentially optimal second-period contract that minimizes transfer payments in that period, and the firm's opportunism, i.e., considering a 'take-the-money-and-run' strategy (sacking first-period transfers and relocating in period 2). This tension can only be resolved by preempting it via a tighter regulation in the first-period. This amounts to a downward-distortion in e_1, that is costly to the regulator. The transfer t_1 required to induce the firm to accept the first-period contract (rather than to relocate immediately) is larger than the total transfer under long-term contracting. This case is depicted in the right panel of Figure 3.1. An implication of Proposition 3.3 is, therefore, that with short-term contracting, the regulator prefers *not* to avert relocation already for lower values of the welfare loss L. In this sense, limited commitment leads to *more* relocation.[24]

Figure 3.2 shows combinations of the parameters π_B and L for which relocation is averted under short-term contracting, in comparison with long-term contracting. As the figure illustrates, the implementation problem that is underlying the results of Proposition 3.3 becomes more severe when the relocation option becomes more attractive (i.e., for larger values of π_B). In contrast, when $\pi_B \leq \pi_B^\sharp$, there is no implementation problem, because offering a contract in period 1 is already sufficient to avert relocation in both periods. If $\pi_B \leq \pi_B^o$ then no transfers are needed to avert relocation.

As a consequence of limited commitment also investments are distorted. In particular, the tougher first-period emission target e_1 leads to an over-investment in abatement capital by the firm.

Corollary 3.2. *Under the optimal sequence of short-term contracts, the implemented investment level is a_A^o for $\pi_B \leq \pi_B^\sharp$ (and $L \geq t^o$), and distorted upwards for $\pi_B > \pi_B^\sharp$ (and $L \geq t^\sharp$).*

Paradoxically, the distortions in e_1 and a can be so severe that the *existence* of an investment opportunity in abatement capital can overall be welfare-reducing. In other words, a seemingly welfare-enhancing investment opportunity, such as investment in abatement capital, may turn out to be welfare-diminishing if it leads to the described conflict of interest between the regulator and the firm. This holds if a higher transfer is required to avert relocation under short-term contracting than in a (hypothetical) situation where $a = 0$ is *exogenously* fixed from the start (and this is common knowledge).

[24] We implicitly assume here that there are several firms that are regulated, and the profit from relocation, π_B, or some other characteristic varies across firms.

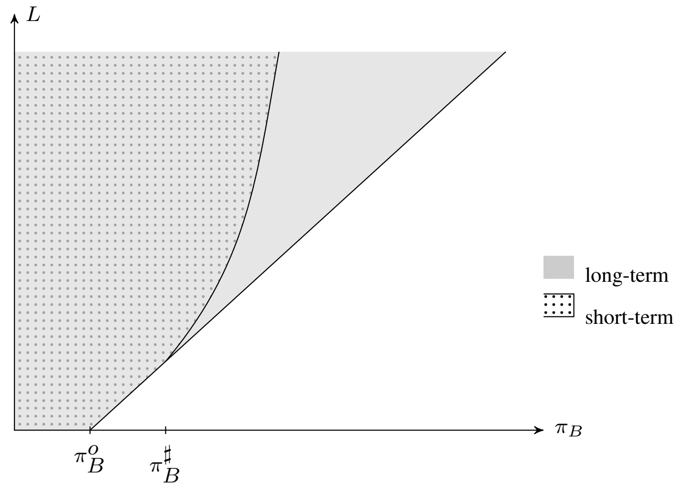

Figure 3.2: (π_B, L) - combinations for which relocation is averted; grey-shaded area: long-term contracting; dotted area: short term.

Corollary 3.3. *If π_B is sufficiently large then $t^\sharp > (1+\delta)(\pi_B - \pi_A^*(0))$, i.e. the regulator would prefer a situation where $a = 0$ is exogenously fixed.*

We close this section by illustrating the above findings in our earlier example.

Example 3.4. *The firm's profit when following location plan 'AB' with first-period emissions e_A^o is given by $V_{AB}(e_A^o) = \frac{5}{2} - \frac{1}{4}\delta^2 + \delta\pi_B$. We have $V_A^o \geq V_{AB}(e_A^o)$ if and only if $\pi_B \leq \pi_B^\sharp = 3 + \frac{3}{4}\delta$. Notice that $a_A^o = 1+\delta$ and hence $\pi_A^*(a_A^o) = 3+\delta > \pi_B$ whenever $\pi_B \leq \pi_B^\sharp$. This demonstrates the lock-in effect, which renders relocation unprofitable even absent any second-period transfer payment. If, however, $\pi_B > \pi_B^\sharp$ a transfer of $t_2 \geq \frac{1}{\delta}\{V_{AB}(e_A^o) - V_A^o\} = \pi_B - \pi_B^\sharp$ is required to implement the long-term contract. Provided the firm indeed chooses investment a_A^o, the sequentially rational second-period transfer is $\max\{0, \pi_B - \pi_A^*(a_A^o)\} = \max\{0, \pi_B - (3+\delta)\}$. Implementation fails, because the latter is strictly lower than $\pi_B - \pi_B^\sharp$, which mirrors the finding of Corollary 3.1. The critical value $e^\sharp$ is given by $e^\sharp = e_A^o - 2(\pi_B - \pi_B^\sharp) = 7 + \frac{\delta}{2} - 2\pi_B$. Consequently, for $\pi_B > \pi_B^\sharp$, the regulator specifies first-period emissions $e_1 = e^\sharp < e_A^o$. The resulting first-period transfer is $t^\sharp = V_B - V_A^o + (\pi_B - \pi_B^\sharp)^2 > V_B - V_A^o$ (if $L \geq t^\sharp$). Investment in this case is $a_A^\sharp = a_A^o + \pi_B - \pi_B^\sharp > a_A^o$.*

To illustrate the finding of Corollary 3.3 notice that $\pi_A^*(0) = 2$. Hence, in the hypothetical situation where investment is impossible the firm earns a maximal per-period profit of 2 and relocation can be averted with a transfer of $\pi_B - 2$ per period. In this case there is no commitment problem, i.e. relocation can be averted permanently *with a total transfer of* $(1+\delta)(\pi_B - 2) = V_B - 2(1+\delta)$. *Obviously, for large* π_B *this expression is smaller than* $t^\sharp$.

3.6 Extensions

In this section we consider extensions of our main model, and analyze to what extent they have an impact on the central result of the previous section, regarding the implementability of outcomes under short-term contracting. First, we consider a situation where the firm's investment is *observable* to the regulator, but remains non-contractible.[25] Second, we focus on a more general objective function of the regulator, that (apart from the firm's location decision) also depends on the firm's emissions, and allows for a benefit to the regulator from averted relocation also in case the firm stays for only one period in A.

3.6.1 Observable investment

Observability of the firm's investment relaxes the implementation problem studied in the previous section to some extent. The reason is, that the regulator can now make the second-period contract offer dependent on the level of investment actually *chosen* by the firm (and not just the anticipated level of a, as in the previous section). As a result, also emission levels $e_1 > e^\sharp$ can now be used to implement SPNE without relocation. Nevertheless, we will show that the optimal long-term contract can only be implemented when $V_A^o \geq V_{AB}(e_A^o)$ (as in the case with an unobservable investment).

Because the regulator now observes the firm's investment level a, the second-period contract entails $e_2 = e^*(a)$ and $t_2 = \max\{0, \pi_B - \pi_A^*(a)\}$, unless the stated t_2 exceeds L (in this case no second-period contract is offered and the firm relocates). Let $\overline{a}$ be the investment level that is just sufficiently large to create a lock-in situation in period 2. Hence, it is implicitly defined by the condition $\pi_A^*(\overline{a}) = \pi_B$.[26] For $a \geq \overline{a}$ no second-period transfer is required to avert relocation and the firm's second-period profit is $\pi_A^*(a)$. Otherwise (for $a < \overline{a}$), the firm is either offered a contract and does not relocate, or there is no second-period contract offer and the firm relocates; in both

[25] Bergemann and Hege (2005) show in a model of project-financing with an infinite time horizon that non-observability of effort may actually be beneficial because it leads to a form of implicit commitment. In our model with a finite horizon, observability is always preferable. Nonetheless, short-term contracting still has severe consequences on implementation.

[26] Existence of $\overline{a}$ follows from Lemma 3.1, result (2).

cases, the firm's profit in period 2 is π_B. Overall, the firm's discounted profit at the investment stage is

$$t_1 + \pi_A(e_1, a) - K(a) + \delta \begin{cases} \pi_A^*(a), & a \geq \bar{a}, \\ \pi_B, & a < \bar{a}. \end{cases} \tag{3.12}$$

After having accepted the first-period contract, the firm chooses its investment to maximize (3.12). The corresponding investment level depends only on e_1. For low values of e_1, namely $e_1 \leq e^\sharp$, the firm invests $a_A(e_1)$. Intuitively, the optimal investment when the firm plans to stay for only one period in country A is, then, already fairly large. The firm then prefers to invest even more, planning to stay also in period 2, even without a second-period transfer. This leads to an optimal investment of $a = a_A(e_1)$. On the other hand, less stringent first-period emission levels $e_1 > e^\sharp$ render large investments unprofitable, so that the firm ends up requiring a transfer in period 2. But in that case its second-period profit is always π_B, so that the firm optimally chooses $a = a_{AB}(e_1)$ even when it does not plan relocate.

Plugging the optimal investment level back into the firm's discounted profit, (3.12), its profit is $t_1 + V_A(e_1)$ whenever $e_1 \leq e^\sharp$, and $t_1 + V_{AB}(e_1)$ whenever $e_1 > e^\sharp$. The first-period transfer that is necessary to implement some first-period emission level e_1 is thus given by $t_1 = V_B - V_A(e_1)$ if $e_1 \leq e^\sharp$, and $t_1 = V_B - V_{AB}(e_1)$ if $e_1 > e^\sharp$. In the latter case, also a positive second-period transfer of $t_2 = \pi_B - \pi_A^*(a_{AB}(e_1))$ is paid. The total (discounted) transfer needed to implement a first-period emission level of $e_1 > e^\sharp$ is $V_B - V_{AB}(e_1) + \delta\big(\pi_B - \pi_A^*(a_{AB}(e_1))\big)$.

Minimizing the total transfer needed to permanently avert relocation leads us to the following result.

Proposition 3.4. *Assume $a_{AB}(e)$ is concave in e.*[27] *With observable investment, the optimal first-period contract is*

- $(t_1, e_1) = (t^o, e_A^o)$, *if* $\pi_B \leq \pi_B^\sharp$ *and* $L \geq t^o$*;*
- $(t_1, e_1) = (t^\sharp, e^\sharp)$, *if* $\pi_B^\sharp < \pi_B \leq \pi_B^{tr}$ *and* $L \geq t^\sharp$*;*
- $(t_1, e_1) = (V_B - V_{AB}(e_A^{tr}), e_A^{tr})$, *if* $\pi_B > \pi_B^{tr}$ *and* $L \geq t^{tr}$*;*

[27] This assumption is sufficient to establish existence and uniqueness of the value e_A^{tr}. Only mild assumptions are required to establish concavity of a_{AB}. E.g., in our illustrative example, $a_{AB}(e)$ is always concave.

- *the null contract otherwise.*

$\pi_B^{tr} > \pi_B^{\sharp}$ *is the critical value for* π_B *for which* $t^{\sharp} = t^{tr}$*. The second-period contract in the third case is* $(t_2, e_2) = (\pi_B - \pi_A^*(a_{AB}(e_A^{tr})), e^*(a_{AB}(e_A^{tr})))$.

Hence, in contrast to the case with unobservable investment, the regulator now has an alternative way to avert relocation, using the possibility to implement a positive second-period transfer. To this end, the regulator adjusts the emissions target in period 1 to the level e_A^{tr}, which induces a sufficiently *small* investment by the firm. In period 2, the regulator then pays a transfer that just averts relocation. However, this option creates a (potential) double inefficiency. Namely, the firm's investment is inefficiently small (given e_1), and in addition the emissions in period 1 are, in general, also distorted.[28] Since the actions implemented by the firm in this case do *not* depend on the value of π_B, whereas the distortions in the case with a lock-in (second case in Proposition 3.4) are increasing in π_B, the regulator implements e_A^{tr} whenever π_B is sufficiently large (larger than π_B^{tr}).

3.6.2 Alternative objective function

In our model as presented so far the regulator's preference only varies in the location of the firm and not directly in the firm's productive choices. Adding a preference over the contractible productive choices of the firm slightly complicates the analysis, but does not reverse the major result of the chapter concerning the implementability of outcomes. In addition, we will also allow for positive benefits of averting relocation only in period 1. We will show that also this modification does not alter the main results. Unlike in the previous subsection, we again assume that a is *not* observable to the regulator.

Suppose, the regulator's payoff can be written as follows:

$$-\chi_1\,(t_1 + D(e_1)) - (1-\chi_1)\,L_1 - \chi_2\,\delta(t_2 + D(e_2)) - (1-\chi_2)\,\delta L_2, \tag{3.13}$$

where $\chi_\tau = 1$ if the firm operates in country A in period τ (and accepts the contract offered in that period), and $\chi_\tau = 0$ otherwise. If the firm relocates in the second period the regulator incurs a loss of L_2 in that period, and if it relocates already in period 1 the regulator incurs an additional loss of $L_1 \geq 0$. Hence, L_1 is the regulator's *benefit*

[28] Whether emissions in period 1 are distorted depends on the specified functions. It turns out that in our illustrative example we have $e_A^{tr} = e_A^o$.

of averting relocation only in period 1. We assume $L_2 \geq L_1$, so that the same payoff structure as in (3.5) is obtained when $L_1 = 0$, while the regulator has an identical interest in averting relocation in each of the two periods when $L_1 = L_2$. $D(e)$ is a penalty function, capturing the domestic damages from the firm's emissions.[29] We assume that $D(e)$ is weakly increasing in e, and that $D(e) = 0$ if $e \leq 0$.

With this payoff structure it is not obvious that the regulator always prefers either immediate relocation or no relocation, because the regulator benefits also from averting relocation only in period 1. However, we argue in the following that due to the sunk costs associated with abatement capital investments, such an outcome is less preferable to either immediate relocation or no relocation and, hence, cannot arise in equilibrium.

Lemma 3.5. *Under the optimal sequence of short-term contracts the firm either relocates immediately or stays for both periods.*

The intuition is straightforward. If the firm stays for one period, it has to receive a transfer that compensates it for not relocating in that period. Because investments are made in the first period, this transfer has to take the investment cost into account. Because these costs are sunk, in period 2 a lower transfer is sufficient to discourage the firm from relocating. This implies that whenever the regulator prefers to avert the firm's relocation in period 1, then he strictly prefers to avert it also in period 2.

Under limited commitment, the regulator thus seeks to find the optimal sequence of short-term contracts that permanently avert relocation with minimal total transfers, taking into consideration also the damages of emissions. If this is too costly, the regulator offers no contract and implements the outcome where the firm relocates immediately.

In the following we derive necessary and sufficient conditions for the implementability of such an outcome, that parallel the results in Section 3.5.2.

To form an equilibrium where the firm does not relocate, the quintuple (t_1, e_1, t_2, e_2, a) again has to satisfy the constraints (PC), (MH-1), and (MH-2). The constraint of sequential optimality now reads as follows

$$(t_2, e_2) \in \arg\min_{\widetilde{t}_2, \widetilde{e}_2} \widetilde{t}_2 + D(\widetilde{e}_2), \quad \text{s.t. } \widetilde{t}_2 + \pi_A(\widetilde{e}_2, a) \geq \max\{\pi_B, \pi_A^*(a)\}. \tag{SO'}$$

[29]When the firm relocates, it may increase its emissions abroad. If pollution is trans-boundary, the regulator will take these emissions into account as well. However, they effectively only raise the fixed welfare loss of relocation and, hence, can be embedded in the parameters L_1 and L_2.

Because the regulator may now prefer a different level of emissions than the firm also in period 2, a further constraint emerges. Namely, the firm should not choose a different investment and thereafter stay in country A also in period 2 without accepting the second-period contract. This leads us to the following additional moral hazard constraint:[30]

$$t_1 + \pi_A(e_1, a) - K(a) + \delta(t_2 + \pi_A(e_2, a)) \geq t_1 + V_A(e_1). \qquad \text{(MH-3)}$$

We can now extend the central result regarding the implementability of outcomes under short-term contracting (see Proposition 3.2) to the generalized payoff structure.

Proposition 3.5. *For a first-period emission level e_1, there exists a second-period contract (t_2, e_2) and an investment level a such that constraints* (MH-1)*,* (MH-2)*,* (SO')*, and* (MH-3) *are satisfied if and only if $V_A(e_1) \geq V_{AB}(e_1)$ and $D'(e^*(a_A(e_1))) = 0$.*

Hence, our result on implementability, which is the central result of this chapter, carries over to the more general payoff function of the regulator. However, the implementation of outcomes becomes even harder. The second condition in Proposition 3.5 requires that given the firm's equilibrium investment a, the regulator's and the firm's interests in the second period are fully aligned. Hence, the regulator must have no incentive to distort the firm's emissions e_2 away from the level that the firm would optimally choose (given a) in the absence of regulation in that period.

The underlying reason for this result is similar as before. Namely, whenever the regulator has an incentive to distort the firm's emissions in period 2, this is anticipated by the firm, and leads to an adjustment in the firm's investment in abatement capital. The regulator, in turn, anticipates this adjustment, and is only willing to compensate the firm for the distortion in second-period emissions, taking this adjustment into account. This shifts the reference point for transfers in the second period, so that the firm is always better off when it plans to reject the second-period contract offer from the start, and invests in abatement capital accordingly (i.e., $a = a_A(e_1)$).[31]

[30] For the sake of brevity we did not write down this constraint under the original payoff structure (see Section 3.5.2), because there it is automatically satisfied given the constraint (SO). This is no longer true under the modified constraint (SO').

[31] This reasoning also applies if the regulator has an incentive to distort the firm's emissions upwards (e.g., in order to trigger a higher choice of output). Anticipating this distortion in the second period, the firm reduces its investment, so that its optimal (un-distorted) emissions are higher in period 2. The

The only way to escape this dilemma is for the regulator to implement an emission level e_1 that preempts the conflict between the regulator's and the firm's interests in period 2. Given the above specification of the regulator's payoff, this holds whenever $e^*(a_A(e_1)) \leq 0$, which implies $D'(e^*(a_A(e_1))) = 0$.[32] Hence, first-period emissions must be set at a sufficiently low level in order to induce a lock in, *and* fulfill the above constraint.[33]

3.7 Conclusion

The chapter identifies a general implementation problem associated with persistent investments by an agent, that yield returns over more than one period. It arises when the principal cannot commit to contractual obligations for the full period of time in which the returns on the investments are incurred. The agent has an outside option, and realizes that in the future, the principal will compensate her only for forgone profits (due to not using the outside option) within a period, and not for her prior investment costs. Hence, the agent is unable to recover the full investment cost, and is better off when she plans to use the outside option in a future period from the start, which implies lower investment costs. We show that the principal is unable to implement outcomes where the agent never uses the outside option *and* requires a strictly positive transfer in a future period. To circumvent this implementation problem, the principal distorts the contract offered to the agent in the first period, where the investment takes place. In particular, by offering more high-powered incentives, the agent is induced to invest more. The outside option, then, becomes less attractive, so that the agent no longer requires a positive transfer in the future and yet refrains from using the outside option.

We frame this general idea in a more specific context. Namely, we analyze the

regulator then only compensates the difference in the firm's profit when choosing its optimal emissions in period 2, given this investment, and the emission level preferred by the regulator.

[32]Depending on the value of the outside option π_B, either the constraint $V_A(e_1) \geq V_{AB}(e_1)$, or the constraint $D'(e^*(a_A(e_1))) = 0$ is binding.

[33]There are other possible modifications of the model that can alleviate the implementation problem. E.g., suppose that in addition to the variable cost of installing an abatement capital stock of a, there is a fixed cost that arises only if a is strictly larger than zero. In that case, the regulator can always induce an investment of zero by setting a sufficiently high emission target for the first period, because this reduces the firm's benefit from investing in abatement capital. But as long as $a = 0$ holds, the local effects from a distortion in the second-period emission target upon the firm's investment vanish. This suggests that – similarly as in the case with an observable investment (see Section 3.6.1) – the regulator has an alternative way to circumvent the implementation problem, by setting a sufficiently loose emission target in the first period.

problem of designing optimal incentive contracts that avert firm relocation. A local regulator aims to avert a firm's relocation in each of two periods. The firm, if staying for at least one period, undertakes some location-specific investment, which is not observable to the regulator. Contracts consist of transfers and targets for an observable productive activity, such as the firm's emissions, output, or employment.

If contracts are long-term, they specify simple subsidy payments, conditional on the firm's location. Optimal long-term contracts do not interfere directly with the firm's operative decisions. This simple structure results because the interests of the regulator and the firm are to some extent aligned. Averting relocation with minimal transfers requires maximal profits of the firm. Therefore, the regulator has no incentive to distort the firm's operative decisions.

With limited commitment an implementation problem arises whenever relocation is sufficiently attractive. Optimal first-period contracts are then more stringent, and implement an inefficiently high investment in order to induce a 'lock-in'. The more attractive the relocation option is, the tougher the contract needs to be, which leads to larger first-period transfers. The distortions that arise due to the implementation problem can be so severe that higher transfers are required to avert the firm's relocation permanently than in a hypothetical situation where the firm cannot invest at all – although a positive investment would be required to avert relocation with minimal transfers.

Our model has an important application in the area of climate policy. When some countries unilaterally introduce prices for emissions, the competitiveness of their energy-intensive industries is harmed. In response, firms may be tempted to relocate to other countries with less stringent environmental regulation. This may be one of the reasons why the EU initially decided to allocate allowances for free in the EU-ETS. Our results indicate that such simple subsidies may not prevent relocation on a *permanent* basis. In order to be effective in this respect, subsidies should be conditioned upon the fulfillment of binding criteria such as firm-specific emission levels, output or employment targets. Such policies are needed whenever policy makers cannot make binding commitments that last for a sufficiently long period of time.

Appendix

3.A Proofs

3.A.1 Proofs of Section 3.4

Proof of Lemma 3.1. Claim (1): $e^*(a)$ is implicitly defined by $\partial\pi_A/\partial e = 0$. By Assumption (A1) this value exists and is unique. Differentiating $\partial\pi_A/\partial e = 0$ w.r.t. a and rearranging yields

$$\frac{\partial e^*}{\partial a} = -\frac{\frac{\partial^2\pi_A}{\partial e\partial a}}{\frac{\partial^2\pi_A}{\partial e^2}} < 0. \tag{3.14}$$

Claim (2): π_A^* is strictly increasing by assumption (A5). To prove concavity of π_A^* differentiate twice, using the envelope-theorem, to get

$$\frac{\partial^2\pi_A^*}{\partial a^2} = \frac{\partial^2\pi_A}{\partial a\partial e}\cdot\frac{\partial e^*}{\partial a} + \frac{\partial^2\pi_A}{\partial a^2}.$$

Using (3.14), this can be written as

$$\frac{\partial^2\pi_A^*}{\partial a^2} = -\frac{\left(\frac{\partial^2\pi_A}{\partial e\partial a}\right)^2}{\frac{\partial^2\pi_A}{\partial e^2}} + \frac{\partial^2\pi_A}{\partial a^2} = \frac{\frac{\partial^2\pi_A}{\partial a^2}\cdot\frac{\partial^2\pi_A}{\partial e^2} - \left(\frac{\partial^2\pi_A}{\partial e\partial a}\right)^2}{\frac{\partial^2\pi_A}{\partial e^2}} \leq 0.$$

The numerator is non-negative by (A3), while the denominator is negative by (A1). Hence the entire expression is negative. Furthermore, (A5) implies $\partial\pi_A^*/\partial a > \varepsilon > 0$ for all a, which yields $\lim_{a\to\infty}\pi_A^*(a) = +\infty$.
Claim (3): $a_A(e)$ is implicitly defined by the first-order condition

$$\frac{\partial\pi_A}{\partial a} - \frac{\partial K}{\partial a} + \delta\frac{\partial\pi_A^*}{\partial a} = 0. \tag{3.15}$$

At $a = 0$ the expression one the left-hand side is strictly positive, by (A2), $K'(0) = 0$, and (A5). Furthermore, boundedness of $\partial\pi_A/\partial a$ by (A2) and strict concavity of K imply that this expression turns negative for large values of a. Existence of $a_A(e)$ then follows from continuity. Furthermore, $\pi_A(e,a) - K(a) + \delta\pi_A^*(a)$ is strictly concave in a, because its components are concave and some even strictly concave, which proves uniqueness of $a_A(e)$. Differentiating (3.15) w.r.t. e and rearranging yields

$$\frac{\partial a_A}{\partial e} = \frac{\frac{\partial^2 \pi_A}{\partial e \partial a}}{\frac{\partial^2 K}{\partial a^2} - \frac{\partial^2 \pi_A}{\partial a^2} - \delta \frac{\partial^2 \pi_A^*}{\partial a^2}} < 0. \tag{3.16}$$

For $a_{AB}(e)$ just repeat the above steps.
Claim (4): By claim (4) both $V_A(e)$ and $V_{AB}(e)$ are well defined. Differentiating $V_A(e)$ twice, using the envelope-theorem, yields

$$\begin{aligned} \frac{\partial^2 V_A}{\partial e^2} &= \frac{\partial^2 \pi_A}{\partial e^2} + \frac{\partial^2 \pi_A}{\partial e \partial a} \cdot \frac{\partial a_A}{\partial e} = \frac{\partial^2 \pi_A}{\partial e^2} + \frac{\left(\frac{\partial^2 \pi_A}{\partial e \partial a}\right)^2}{\frac{\partial^2 K}{\partial a^2} - \frac{\partial^2 \pi_A}{\partial a^2} - \delta \frac{\partial^2 \pi_A^*}{\partial a^2}} \\ &= \frac{\frac{\partial^2 K}{\partial a^2} \cdot \frac{\partial^2 \pi_A}{\partial e^2} - \left[\frac{\partial^2 \pi_A}{\partial a^2} \cdot \frac{\partial^2 \pi_A}{\partial e^2} - \left(\frac{\partial^2 \pi_A}{\partial e \partial a}\right)^2\right] - \delta \frac{\partial^2 \pi_A^*}{\partial a^2} \cdot \frac{\partial^2 \pi_A}{\partial e^2}}{\frac{\partial^2 K}{\partial a^2} - \frac{\partial^2 \pi_A}{\partial a^2} - \delta \frac{\partial^2 \pi_A^*}{\partial a^2}} < 0. \end{aligned}$$

Concavity of $V_{AB}(e)$ is proven in the same way (not shown). Using the envelope-theorem, the first-order condition for maximizing $V_A(e)$ is $\frac{\partial \pi_A}{\partial e}(e, a_A(e)) = 0$. By (A1) and continuity there exits some value e that satisfies this equation. Uniqueness follows from strict concavity of $V_A(e)$. Similarly, maximizing $V_{AB}(e)$ yields the first-order condition $\frac{\partial \pi_A}{\partial e}(e, a_{AB}(e)) = 0$, existence and uniqueness follow as before.
Claim (5): $a_{AB}(e)$ is defined by the first-order condition

$$\frac{\partial \pi_A}{\partial a} - \frac{\partial K}{\partial a} = 0. \tag{3.17}$$

Comparing this to (3.15), noticing that π_A^* is strictly increasing and by concavity of the respective objectives, we find that $a_A(e) > a_{AB}(e)$ for all e. □

Proof of Lemma 3.2. Assume $V_{AB}(e_1) \geq V_B$, which can be written as

$$V_{AB}(e_1) = \pi_A(e_1, a_{AB}(e_1)) - K(a_{AB}(e_1)) + \delta\pi_B \geq \pi_B + \delta\pi_B = V_B.$$

But this implies $\pi_A(e_1, a_{AB}(e_1)) > \pi_B$ and therefore

$$\begin{aligned} V_A(e_1) &= \max_a \ \pi_A(e_1, a) - K(a) + \delta\pi_A^*(a) \\ &\geq \pi_A(e_1, a_{AB}(e_1)) - K(a_{AB}(e_1)) + \delta\pi_A(e_1, a_{AB}(e_1)) \\ &> \pi_A(e_1, a_{AB}(e_1)) - K(a_{AB}(e_1)) + \delta\pi_B \\ &= V_{AB}(e_1). \end{aligned}$$

This proves our claim. □

Proof of Lemma 3.3. As is discussed in the main text, the optimal profit from not relocating is V_A^o. The profit from immediate relocation is V_B. As a consequence of Lemma 3.2 we have $V_{AB}(e_1) < \max\{V_A^o, V_B\}$ for all e_1. Therefore, the firm prefers immediate relocation whenever $V_B > V_A^o$ and no relocation otherwise. Solving $V_A^o = V_B$ for π_B leads to the definition of π_B^o. □

3.A.2 Proofs of Section 3.5

Proof of Proposition 3.1. As is argued in the main text, the regulator's problem is to minimize (3.11) over e_1 and e_2. This is equivalent to maximizing $\pi_A(e_1, a) - K(a) + \delta\pi_A(e_2, a)$ over a, e_1, and e_2. Maximizing first over e_2 and a yields $V_A(e_1)$. Maximizing this over e_1 yields $e_1 = e_A^o$. By comparing the respective first-order conditions we get $e_2 = e_1$. The total transfer required is $t^o = V_B - V_A^o$. The regulator offers this contract whenever $t^o \leq L$. □

Proof of Proposition 3.2. When (SO) is satisfied, the firm's second-period profit is $t_2 + \pi_A^*(a)$. By the envelope-theorem, (MH-1) then implies that the firm's total profit is $t_1 + \delta t_2 + V_A(e_1)$. This justifies constraint (MH-2′), as a replacement for (MH-2). Now first assume $V_A(e_1) \geq V_{AB}(e_1)$, which can be stated as

$$\max_a \ \pi_A(e_1, a) - K(a) + \delta\pi_A^*(a) \geq \max_a \ \pi_A(e_1, a) - K(a) + \delta\pi_B. \tag{3.18}$$

This implies $\pi_A^*(a_A(e_1)) > \pi_B$, where $a_A(e_1)$ denotes the maximizer of the left-hand side. Hence, the second-period contract $(t_2, e_2) = (0, e^*(a_A(e_1)))$ satisfies (SO), given $a = a_A(e_1)$. By construction, (MH-1) and (MH-2) are satisfied, given (t_2, e_2).

Next assume $V_A(e_1) < V_{AB}(e_1)$. Constraints (MH-1) and (SO) imply $a = a_A(e_1)$ and the second-period contract offer entails $t_2 = \max\{0, \pi_B - \pi_A^*(a_A(e_1))\}$ and $e_2 = e^*(a_A(e_1))$. As indicated above, (MH-2) can be replaced by (MH-2′). Therefore, necessary for all three constraints to hold is $\delta t_2 \geq V_{AB}(e_1) - V_A(e_1) > 0$. Further, note that

$$\begin{aligned} \delta t_2 &\geq V_{AB}(e_1) - V_A(e_1) \\ &= \max_a \big\{\pi_A(e_1, a) - K(a) + \delta\pi_B\big\} - \max_a \big\{\pi_A(e_1, a) - K(a) + \delta\pi_A^*(a)\big\} \\ &> \delta\big(\pi_B - \pi_A^*(a_A(e_1))\big). \end{aligned}$$

Therefore $t_2 > \pi_B - \pi_A^*(a_A)$ and together with $t_2 > 0$, as shown above, we get $t_2 > \max\{0, \pi_B - \pi_A^*(a_A)\}$ – this contradicts (SO). □

Proof of Corollary 3.1. The result on implementability follows from Proposition 3.2. Regarding $\pi_B^\sharp$ notice that $V_A^o > V_{AB}(e_A^o)$ for $\pi_B = \pi_B^o$ by Lemma 3.2. Because $V_{AB}(e_A^o)$ strictly increases with π_B, while V_A^o is independent of π_B, we get $\pi_B^\sharp > \pi_B^o$. □

Proof of Lemma 3.4. By the envelope-theorem $\partial V_A/\partial\pi_B = 0 < \delta = \partial V_{AB}/\partial\pi_B$. Furthermore, using $a_A(e) > a_{AB}(e)$, it holds that

$$\frac{\partial V_A}{\partial e} = \frac{\partial \pi_A}{\partial e}(e, a_A(e)) < \frac{\partial \pi_A}{\partial e}(e, a_{AB}(e)) = \frac{\partial V_{AB}}{\partial e}. \tag{3.19}$$

Together with $V_A(e_A^o) = V_{AB}(e_A^o)$ for $\pi_B = \pi_B^\sharp$ (from Corollary 3.1) this yields $e^\sharp < e_A^o$ and $e^\sharp$ strictly decreases with π_B.
It remains to prove that $e^\sharp > \underline{e}$ for all π_B. To see this, notice that $V_A(e) = V_{AB}(e)$ at $\delta = 0$ for all e. Also, $\partial V_A/\partial\delta = \pi_A^*(a)$ and $\partial V_{AB}/\partial\delta = \pi_B$. For $e \to \underline{e}$ we have by (A1) and strictly convex K that $a_A(e) \to \infty$. As this holds irrespective of δ, we have that $V_A(e) > V_{AB}(e)$ for $e \to \underline{e}$ which completes the proof. □

Proof of Proposition 3.3. We determine the cost of implementing an equilibrium with no relocation. Recall from the proof of Proposition 3.2 that there is no second-period transfer. As long as $\pi_B \leq \pi_B^\sharp$, by Corollary 3.1, e_A^o is implementable and minimizes the cost over the set of implementable first-period emission levels; the required (total) transfer is $t^o = V_B - V_A^o$. If $\pi_B > \pi_B^\sharp$, we have $e_A^o > e^\sharp$. Therefore, the regulator cannot use e_A^o to implement an outcome with no relocation. By the concavity of V_A, implementing $e^\sharp$ requires the smallest transfer, which is equal to $t^\sharp = V_B - V_A(e^\sharp)$. □

Proof of Corollary 3.2. Trivial for $\pi_B \leq \pi_B^\sharp$. For $\pi_B > \pi_B^\sharp$ recall that $a_A(e)$ decreases in e (Lemma 3.1), and $e^\sharp < e_A^o$. The result follows. □

Proof of Corollary 3.3. Recall that $t^\sharp = V_B - V_A(e^\sharp)$. On the other hand, the transfer to avert relocation when $a = 0$ is given by $t^{a=0} = V_B - (1+\delta)\pi_A^*(0)$, and no implementation problem arises in this case as a is fixed. Therefore $t^\sharp - t^{a=0} = -V_A(e^\sharp) + (1+\delta)\pi_A^*(0)$. Now, from Lemma 3.4 we have $e^\sharp \to \underline{e}$ for $\pi_B \to \infty$ and by strict concavity of V_A this implies $V_A(e^\sharp) \to -\infty$. Consequently, $t^\sharp - t^{a=0} \to \infty$, which proves the claim. □

3.A.3 Proofs of Section 3.6

Proof of Proposition 3.4. We first characterize the firm's optimal investment decision, i.e. the maximizer of (3.12). We distinguish three cases:

i) $\bar{a} \leq a_{AB}(e_1)$. By concavity of $\pi_A(e,a) - K(a) + \delta\pi_B$ (see the proof of Lemma 3.1), we have for all $a \leq \overline{a}$:

$$\begin{aligned}\pi_A(e_1, a) - K(a) + \delta\pi_B &\leq \pi_A(e_1, \overline{a}) - K(\overline{a}) + \delta\pi_B \\ &= \pi_A(e_1, \overline{a}) - K(\overline{a}) + \delta\pi_A^*(\overline{a}).\end{aligned}$$

Furthermore, because $\bar{a} \leq a_{AB}(e_1) < a_A(e_1)$, we have $V_A(e_1) \geq \pi_A(e_1, a) - K(a) + \delta\pi_A^*(a)$ for all $a \geq \bar{a}$. Consequently, $a = a_A(e_1)$ maximizes the firm's profit in this case and this maximal profit is $V_A(e_1)$.

ii) $a_A(e_1) \leq \bar{a}$. Similar to the previous case we have for all $a \geq \overline{a}$:

$$\begin{aligned}\pi_A(e_1, a) - K(a) + \delta\pi_A^*(a) &\leq \pi_A(e_1, \overline{a}) - K(\overline{a}) + \delta\pi_A^*(\overline{a}) \\ &= \pi_A(e_1, \overline{a}) - K(\overline{a}) + \delta\pi_B.\end{aligned}$$

Furthermore, because $a_{AB}(e_1) < a_A(e_1) \leq \bar{a}$, we have $V_{AB}(e_1) \geq \pi_A(e_1, a) - K(a) + \delta\pi_B$ for all $a \leq \bar{a}$. Consequently, $a = a_{AB}(e_1)$ maximizes the firm's expected profit in this case and this maximal profit is $V_{AB}(e_1)$.

iii) $a_{AB}(e_1) < \bar{a} < a_A(e_1)$. By the above arguments the firm's profit has two local maxima: at $a = a_A(e_1)$ and at $a = a_{AB}(e_1)$, such that the maximal profit is either $V_A(e_1)$ or $V_{AB}(e_1)$. Because $V_A(e) > V_{AB}(e)$ holds if and only if $e < e^\sharp$, we find that the firm's maximal profit, given $a_{AB}(e_1) < \bar{a} < a_A(e_1)$, is thus $V_A(e_1)$ if $e_1 \leq e^\sharp$, and $V_{AB}(e_1)$ if $e_1 > e^\sharp$.

Therefore, the firm's profit after having accepted a first-period contract offer (t_1, e_1) is

$$t_1 + \begin{cases} V_A(e_1), & e_1 \leq e^\sharp, \\ V_{AB}(e_1), & e_1 > e^\sharp. \end{cases} \tag{3.20}$$

We here implicitly assume that the firm always chooses $a_A(e_1)$ when $e_1 = e^\sharp$, although it is indifferent. This is without loss of generality, because the regulator chooses the equilibrium, in case there are multiple, and it is obvious that the first-period transfer to implement $e_1 = e^\sharp$ is unaffected by the continuation, but in case the firm chooses $a_{AB}(e^\sharp)$ the regulator has to pay a strictly positive second-period transfer to avert relocation in period 2.
The total transfer to avert relocation is given by

$$t(e_1) = \begin{cases} V_B - V_A(e_1), & e_1 \leq e^\sharp, \\ V_B - V_{AB}(e_1) + \delta\big(\pi_B - \pi_A^*(a_{AB}(e_1))\big), & e_1 > e^\sharp. \end{cases} \tag{3.21}$$

In case $e_1 \leq e^\sharp$ this is trivial, because it implies $a = a_A(e_1) > \overline{a}$ and therefore a first-period transfer is sufficient (this already follows from Lemma 3.4). Now consider $e_1 > e^\sharp$, and suppose $\pi_A^*(a_{AB}(e_1)) \geq \pi_B$. This would imply

$$\begin{aligned} V_{AB}(e_1) &= \pi_A(e_1, a_{AB}(e_1)) - K(a_{AB}(e_1)) + \delta\pi_B \\ &\leq \pi_A(e_1, a_{AB}(e_1)) - K(a_{AB}(e_1)) + \delta\pi_A^*(a_{AB}(e_1)) < V_A(e_1), \end{aligned}$$

which yields $e_1 < e^\sharp$ – a contradiction. Thus, $\pi_A^*(a_{AB}(e_1)) < \pi_B$, so that the minimal second-period transfer required to implement an outcome with no relocation is $t_2 = \pi_B - \pi_A^*(a_{AB}(e_1))$.
The regulator now chooses e_1 in order to minimize (3.21). The first case ($\pi_B \leq \pi_B^\sharp \Leftrightarrow e_A^o \leq e^\sharp$) follows readily from Corollary 3.1. For the remainder, assume $e_A^o > e^\sharp$, i.e.

$\pi_B > \pi_B^\sharp$. By strict concavity of $V_A(e)$ we have

$$t(e_1) = V_B - V_A(e_1) > V_B - V_A(e^\sharp) = t^\sharp \qquad \forall e_1 < e^\sharp.$$

So it cannot be optimal to implement some $e_1 < e^\sharp$. For $e_1 > e^\sharp$, the required transfer is $\widetilde{t}(e_1) = V_B - V_{AB}(e_1) + \delta\big(\pi_B - \pi_A^*(a_{AB}(e_1))\big)$. Denote e_A^{tr} the minimizer of $\widetilde{t}(e_1)$. By Lemma 3.1, the function $V_{AB}(e_1)$ is strictly concave. Furthermore, because π_A^* is concave and strictly increasing by Lemma 3.1, the composition with the concave function $a_{AB}(e_1)$ is also concave. Therefore, $\widetilde{t}(e_1)$ is strictly convex for *all* $e_1 \in (\underline{e}, \overline{e})$. Furthermore, by (A1) and Lemma 3.1 the minimizer is interior, i.e. $e_A^{\mathrm{tr}} \in (\underline{e}, \overline{e})$ exists. Now suppose $e_A^{\mathrm{tr}} \leq e^\sharp$. Then $t(e^\sharp) \geq \widetilde{t}(e^\sharp) > \widetilde{t}(e_1)$ for all $e_1 > e^\sharp$ so that $e_1 = e^\sharp$ leads to minimal (total) transfers. Hence the relevant cases are where $e_A^{\mathrm{tr}} > e^\sharp$. Notice, that $\widetilde{t}(e_A^{\mathrm{tr}})$ does not depend on π_B, and that for $\pi_B = \pi_B^\sharp$ we have $V_A^{\mathrm{tr}}(e_A^{\mathrm{tr}}) < V_A(e^\sharp)$. Because $t(e^\sharp)$ strictly increases with π_B and converges to $+\infty$, there exists a level π_B^{tr} such that $t(e_A^{\mathrm{tr}}) < t(e^\sharp)$ if and only if $\pi_B > \pi_B^{\mathrm{tr}}$. This completes the proof. □

Proof of Lemma 3.5. Suppose the regulator offers (t_1, e_1) in the first period, which is accepted by the firm and relocation in period 2 occurs. Denote $\hat{a}$ the equilibrium value of the firm's investment. Because the firm relocates in period 2, we must have $\pi_A^*(\hat{a}) \leq \pi_B$. Regarding the first-period transfer, it has to hold that $t_1 \geq V_B - V_{AB}(e_1)$, in order to be accepted by the firm. Furthermore, we must have $L_1 \geq t_1 + D(e_1)$, otherwise the regulator prefers not to offer the contract at all. But then we have

$$\begin{aligned} 0 \leq\ & L_1 - t_1 - D(e_1) \ \leq\ L_1 - V_B + V_{AB}(e_1) - D(e_1) \\ =\ & L_1 - \pi_B + \pi_A(e_1, \hat{a}) - K(\hat{a}) - D(e_1) \ <\ L_2 - \pi_B + \pi_A(e_1, \hat{a}) - D(e_1). \end{aligned}$$

Now, because $\pi_A^*(\hat{a}) \leq \pi_B$, the optimal contract to keep the firm in country A in period 2 is the solution to

$$\min_{t_2, e_2} t_2 + D(e_2) \quad \text{s.t.}\ \ t_2 + \pi_A(e_2, \hat{a}) \geq \pi_B. \tag{3.22}$$

Clearly, the solution to this is $e_2 = \arg\max_e \pi_A(e, \hat{a}) - D(e)$ and $t_2 = \pi_B - \pi_A(e_2, \hat{a})$.

Together with the above, the regulator's benefit from offering this contract is

$$\begin{aligned} L_2 - t_2 - D(e_2) &= L_2 - \pi_B + \pi_A(e_2, \hat{a}) - D(e_2) \\ &> L_2 - \pi_B + \pi_A(e_1, \hat{a}) - D(e_1) > 0, \end{aligned}$$

where the first inequality holds because e_2 maximizes $\pi_A(e, \hat{a}) - D(e)$, and the second inequality was shown above to hold. Hence, the regulator strictly prefers offering a contract in period 2 that averts relocation.
Notice that the method of proof also rules out random relocation in period 2. Hence, either immediate relocation or no relocation can be optimal. □

Proof of Proposition 3.5. Let (t_1, e_1, t_2, e_2, a) be the outcome to be implemented.
Assume first that $V_A(e_1) \geq V_{AB}(e_1)$ and the second-period contract entails $e_2 \neq e^*(a)$. Because $D' \geq 0$ this implies $e_2 < e^*(a)$ and thus (MH-1) implies $a > a_A(e_1)$. But then $\pi_A^*(a) > \pi_A^*(a_A(e_1)) > \pi_B$. The firm's second-period profit, including the transfer $t_2 = \pi_A^*(a) - \pi_A(e_2, a)$, is therefore $\pi_A^*(a)$, but then (MH-3) is clearly violated because $a \neq a_A(e_1)$ is not the maximizer of $\pi_A(e_1, \widetilde{a}) - K(\widetilde{a}) + \delta\pi_A^*(\widetilde{a})$.
Next assume $V_A(e_1) \geq V_{AB}(e_1)$ and the second-period contract entails $e_2 = e^*(a)$. Then (MH-3) is trivially satisfied. Also (MH-2) holds, by the arguments used in proving Lemma 3.2. Constraint (SO') is only satisfied when the regulator indeed prefers to keep the firm without distorting its second-period emissions, for which the second condition from the Proposition is both necessary and sufficient.
Lastly, assume $V_A(e_1) < V_{AB}(e_1)$. If $\pi_A^*(a) \geq \pi_B$ the firm's equilibrium payoff is $t_1 + \pi_A(e_1, a) - K(a) + \delta\pi_A^*(a) \leq t_1 + V_A(e_1) < t_1 + V_{AB}(e_1)$, hence (MH-2) is violated. If on the other hand $\pi_A^*(a) < \pi_B$ the firm's equilibrium payoff is $t_1 + \pi_A(e_1, a) - K(a) + \delta\pi_B$. Because $D' \geq 0$ we must have $e_2 \leq e^*(a)$ and, therefore, $\partial\pi_A/\partial a \mid_{e_2,a} > 0$ by assumptions (A1) and (A5), which implies $a \neq a_{AB}(e_1)$. Consequently (MH-2) is violated because a is not the maximizer of $\pi_A(e_1, \widetilde{a}) - K(\widetilde{a}) + \delta\pi_B$. □

3.B Restriction to pure strategies

Here we argue that allowing for mixed strategies does not soften the regulator's implementation problem identified in Proposition 3.2. An equilibrium in mixed strategies is characterized by a randomized strategy of the firm, i.e. a distribution on a subset A of the real line, and a mechanism that the regulator offers in period 2. By the revelation principle, the latter mechanism can be assumed to be direct, incentive compatible and truth-telling.[34]

For simplicity we focus in our analysis on the discrete case, i.e. where the firm randomizes over the discrete set of investment levels $\mathcal{A} = \{a^1, \ldots, a^n\}$. Clearly, there must exist $\hat{a} \in \mathcal{A}$ which receives no positive rent. Denote the contract this types accepts in equilibrium as $(\hat{t}_2, \hat{e}_2)$. Then it must hold that

$$\hat{t}_2 + \pi_A(\hat{e}_2, \hat{a}) = \pi_B. \tag{3.23}$$

Now consider the firm's investment choice. First of all, $\hat{a}$ must maximize the following expression

$$t_1 + \pi_A(e_1, a) - K(a) + \delta(\hat{t}_2 + \pi_A(\hat{e}_2, a)). \tag{3.24}$$

Second, because of (3.23), $\hat{a}$ also maximizes

$$t_1 + \pi_A(e_1, a) - K(a) + \delta\pi_B. \tag{3.25}$$

Using the first order-conditions for (3.24) and (3.25), $\hat{a}$ has to satisfy

$$\frac{\partial \pi_A}{\partial a}(\hat{e}_2, \hat{a}) = 0. \tag{3.26}$$

Because the function π_A is strictly concave in a for any value e, we conclude that

$$\pi_A(\hat{e}_2, a) < \pi_A(\hat{e}_2, \hat{a}), \quad \forall a \neq \hat{a}. \tag{3.27}$$

Together with (3.23) this implies

$$\hat{t}_2 + \pi_A(\hat{e}_2, a) < \pi_B, \quad \forall a \in \mathcal{A} \setminus \{\hat{a}\}. \tag{3.28}$$

Thus, no other type has the incentive to mimic type $\hat{a}$, because any type is guaranteed a profit of at least π_B. But this implies that there exists a second type $a' \neq \hat{a}$ that also receives no rent, because otherwise we could reduce all transfers to types $a \neq \hat{a}$

[34]Because only allocations matter for providing investment incentives to the firm, replacing an arbitrary mechanism that leads to a particular allocation with its direct and incentive compatible counterpart is indeed without loss of generality.

without violating any incentive constraint. This type a' also has to maximize (3.25). Because (3.25) has a unique maximizer, namely $a_{AB}(e_1)$, this leads to a contradiction.

Bibliography

ALBANO, G., AND A. LIZZERI (2001): "Strategic certification and provision of quality," *International Economic Review*, 42(1), 267–283.

ATTAR, A., AND E. CAMPIONI (2003): "Costly state verification and debt contracts: a critical resume," *Research in Economics*, 57(4), 315–343.

BABIKER, M. H. (2005): "Climate change policy, market structure, and carbon leakage," *Journal of International Economics*, 65(2), 421–445.

BARON, D. P. (1984): "Regulatory Strategies Under Asymmetric Information," in *Bayesian Models in Economic Theory*, ed. by M. Boyer, and R. Kihlstrom. North-Holland.

——— (1989): "Optimal Regulatory Mechanism Design," in *Handbook of Industrial Organization*, ed. by R. Schmalensee, and R. Willig. North-Holland.

BARON, D. P., AND D. BESANKO (1984): "Regulation, Asymmetric Information, and Auditing," *RAND Journal of Economics*, 15(4), 447–470.

BARON, D. P., AND R. B. MYERSON (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50(4), 911–930.

BERGEMANN, D., AND U. HEGE (1998): "Venture capital financing, moral hazard, and learning," *Journal of Banking & Finance*, 22(6), 703–735.

——— (2005): "The financing of innovation: learning and stopping," *RAND Journal of Economics*, 36(4), 719–752.

BESTER, H., AND R. STRAUSZ (2001): "Contracting with imperfect commitment and the revelation principle: the single agent case," *Econometrica*, 69(4), 1077–1098.

——— (2007): "Contracting with imperfect commitment and noisy communication," *Journal of Economic Theory*, 136(1), 236–259.

BHASKAR, V. (2014): "The Ratchet Effect Re-examined: A Learning Perspective," CEPR London. http://www.cepr.org/pubs/new-dps/dplist.php?dpno=109956.

BIGLAISER, G. (1993): "Middlemen as Experts," *RAND Journal of Economics*, 24(2), 212–223.

BLUME, A., O. J. BOARD, AND K. KAWAMURA (2007): "Noisy talk," *Theoretical Economics*, 2(4), 395–440.

BOLTON, P., X. FREIXAS, AND J. SHAPIRO (2012): "The Credit Ratings Game," *Journal of Finance*, 67(1), 85–112.

BORDER, K. C., AND J. SOBEL (1987): "Samurai accountant: A theory of auditing and plunder," *Review of Economic Studies*, 54(4), 525–540.

BUCOVETSKY, S. (2005): "Public input competition," *Journal of Public Economics*, 89(9), 1763–1787.

CHATTERJEE, K., S. MORTON, AND A. MUKHERJI (2008): "Strategic audit policies without commitment," in *Pareto Optimality, Game Theory And Equilibria*, ed. by P. Pardalos, A. Migdalas, and L. Pitsoulis, pp. 407–436. Springer.

CHE, Y.-K., AND J. SÁKOVICS (2004): "A dynamic theory of holdup," *Econometrica*, 72(4), 1063–1103.

CHIAPPORI, P.-A., I. MACHO, P. REY, AND B. SALANIÉ (1994): "Repeated moral hazard: The role of memory, commitment, and the access to credit markets," *European Economic Review*, 38(8), 1527–1553.

CRAWFORD, V. P., AND J. SOBEL (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

DEMSKI, J. S., D. E. SAPPINGTON, AND P. T. SPILLER (1987): "Managing Supplier Switching," *RAND Journal of Economics*, 18(1), 77–97.

DUBEY, P., AND J. GEANAKOPLOS (2010): "Grading exams: 100, 99,..., 1 or A, B, C? Incentives in Games of Status," *Games and Economic Behavior*, 69(1), 72–94.

DUNNE, S. A., AND M. A. LOEWENSTEIN (1995): “Costly Verification of Cost Performance and the Competition for Incentive Contracts,” *RAND Journal of Economics*, 26(4), 690–703.

FARHI, E., J. LERNER, AND J. TIROLE (2013): “Fear of rejection? Tiered certification and transparency,” *The RAND Journal of Economics*, 44(4), 610–631.

FAURE-GRIMAUD, A., E. PEYRACHE, AND L. QUESADA (2009): “The ownership of ratings,” *RAND Journal of Economics*, 40(2), 234–257.

FORGES, F. (1986): “An Approach To Communication Equilibria,” *Econometrica*, 54(6), 1375–1385.

FORGES, F. (1990): “Equilibria with Communication in a Job Market Example,” *The Quarterly Journal of Economics*, 105(2), 375–98.

FREIXAS, X., R. GUESNERIE, AND J. TIROLE (1985): “Planning under incomplete information and the ratchet effect,” *Review of Economic Studies*, 52(2), 173–191.

FUDENBERG, D., B. HOLMSTROM, AND P. MILGROM (1990): “Short-term contracts and long-term agency relationships,” *Journal of Economic Theory*, 51(1), 1–31.

GIBBONS, R. (1986): “Piece-Rate Incentive Schemes,” *Journal of Labor Economics*, 5(4), 413–29.

GOLTSMAN, M., J. HÖRNER, G. PAVLOV, AND F. SQUINTANI (2009): “Mediation, arbitration and negotiation,” *Journal of Economic Theory*, 144(4), 1397–1420.

GRAETZ, M. J., J. F. REINGANUM, AND L. L. WILDE (1986): “The tax compliance game: Toward an interactive theory of law enforcement,” *Journal of Law, Economics, & Organization*, 2(1), 1–32.

HARSTAD, B. (2012): “Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations,” *Review of Economic Studies*, 79(4), 1527–1557.

HART, O. (1995): *Firms, contracts, and financial structure*. Oxford University Press.

HART, O., AND J. MOORE (1988): “Incomplete Contracts and Renegotiation,” *Econometrica*, 56(4), 755–785.

HAUFLER, A., AND I. WOOTON (2010): “Competition for firms in an oligopolistic industry: The impact of economic integration,” *Journal of International Economics*, 80(2), 239–248.

HORSTMANN, I. J., AND J. R. MARKUSEN (1992): “Endogenous market structures in international trade (natura facit saltum),” *Journal of International Economics*, 32(1), 109–129.

JOSKOW, P. L. (1987): “Contract duration and relationship-specific investments: Empirical evidence from coal markets,” *American Economic Review*, 77(1), 168–185.

KARTASHEVA, A., AND B. YILMAZ (2013): “Precision of Ratings,” http://ssrn.com/abstract=2232998.

KHALIL, F. (1997): “Auditing Without Commitment,” *RAND Journal of Economics*, 28(4), 629–640.

KHALIL, F., AND J. LAWARRÉE (1995): “Collusive auditors,” *American Economic Review*, 85(2), 442–446.

——— (2006): “Incentives for Courruptible Auditors in the Absence of Commitment,” *Journal of Industrial Economics*, 54(2), 269–291.

KHALIL, F., AND B. M. PARIGI (1998): “Loan Size as a Commitment Device,” *International Economic Review*, 39(1), 135–50.

KOFMAN, F., AND J. LAWARRÉE (1993): “Collusion in Hierarchical Agency,” *Econometrica*, 61(3), 629–56.

KREPS, D. M., AND R. WILSON (1982): “Reputation and imperfect information,” *Journal of Economic Theory*, 27(2), 253–279.

KRISHNA, V., AND J. MORGAN (2004): “The art of conversation: eliciting information from experts through multi-stage communication,” *Journal of Economic Theory*, 117(2), 147–179.

LAFFONT, J.-J., AND J. TIROLE (1988): “The Dynamics of Incentive Contracts,” *Econometrica*, 56(5), 1153–75.

LANGINIER, C., AND B. A. BABCOCK (2008): "Agricultural production clubs: Viability and welfare implications," *Journal of Agricultural & Food Industrial Organization*, 6(1).

LAZEAR, E. P. (1986): "Salaries and Piece Rates," *Journal of Business*, 59(3), 405–431.

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

LIZZERI, A. (1999): "Information Revelation and Certification Intermediaries," *RAND Journal of Economics*, 30(2), 214–231.

MAILATH, G., AND L. SAMUELSON (2012): *Repeated Games and Reputations: Long-run Relationships*. Oxford University Press.

MANSO, G. (2011): "Motivating innovation," *Journal of Finance*, 66(5), 1823–1860.

MARKUSEN, J. R., E. R. MOREY, AND N. OLEWILER (1993): "Environmental policy when market structure and plant locations are endogenous," *Journal of Environmental Economics and Management*, 24(1), 69–86.

——— (1995): "Competition in regional environmental policies when plant locations are endogenous," *Journal of Public Economics*, 56(1), 55–77.

MASON, C. F. (2011): "Eco-labeling and market equilibria with noisy certification tests," *Environmental and Resource Economics*, 48(4), 537–560.

MATHIS, J., J. MCANDREWS, AND J. ROCHET (2009): "Rating the raters: are reputation concerns powerful enough to discipline rating agencies?," *Journal of Monetary Economics*, 56(5), 657–674.

MELUMAD, N. D., AND D. MOOKHERJEE (1989): "Delegation as commitment: the case of income tax, audits," *RAND Journal of Economics*, 20(2), 139–163.

MITUSCH, K., AND R. STRAUSZ (2005): "Mediation in situations of conflict and limited commitment," *Journal of Law, Economics, and Organization*, 21(2), 467–500.

MOOKHERJEE, D., AND I. PNG (1989): "Optimal Auditing, Insurance, and Redistribution," *Quarterly Journal of Economics*, 104(2), 399–415.

MOTTA, M., AND J.-F. THISSE (1994): "Does environmental dumping lead to delocation?," *European Economic Review*, 38(3), 563–576.

MYERSON, R. B. (1982): "Optimal coordination mechanisms in generalized principal–agent problems," *Journal of Mathematical Economics*, 10(1), 67–81.

——— (1986): "Multistage Games with Communication," *Econometrica*, 54(2), 323–358.

PAGANO, M., AND P. VOLPIN (2012): "Securitization, transparency, and liquidity," *Review of Financial Studies*, 25(8), 2417–2453.

PICARD, P. (1996): "Auditing claims in the insurance market with fraud: The credibility issue," *Journal of Public Economics*, 63(1), 27–56.

POLLRICH, M., AND R. C. SCHMIDT (2014): "Optimal incentive contracts to avert firm relocation," SFB/TR 15 Discussion Paper No. 480, available at http://www.sfbtr15.de/uploads/media/480.pdf.

POLLRICH, M., AND L. WAGNER (2014): "Informational opacity and honest certification," SFB/TR 15 Discussion Paper No. 481, available at http://www.sfbtr15.de/uploads/media/481.pdf.

RAHMAN, D., AND I. OBARA (2010): "Mediated partnerships," *Econometrica*, 78(1), 285–308.

REY, P., AND B. SALANIÉ (1990): "Long-term, Short-term and Renegotiation: On the Value of Commitment in Contracting," *Econometrica*, 58(3), 597–619.

SCHMIDT, R. C., AND J. HEITZIG (2014): "Carbon leakage: Grandfathering as an incentive device to avert firm relocation," *Journal of Environmental Economics and Management*, 67(2), 209–233.

SHAPIRO, C. (1983): "Premiums for high quality products as returns to reputations," *Quarterly Journal of Economics*, 98(4), 659.

SKRETA, V. (2006): "Sequentially optimal mechanisms," *The Review of Economic Studies*, 73(4), 1085–1111.

——— (2013): "Optimal auction design under non-commitment," NYU Working Paper No. 2451/31787. http://ssrn.com/abstract=1281906.

STRAUSZ, R. (2005): “Honest certification and the threat of capture,” *International Journal of Industrial Organization*, 23(1-2), 45–62.

——— (2012): “Mediated contracts and mechanism design,” *Journal of Economic Theory*, 147(3), 1280–1290.

TIROLE, J. (1986): “Hierarchies and Bureaucracies: On the Role of Collusion in Organizations,” *Journal of Law, Economics, and Organization*, 2(2), 181–214.

ULPH, A. (1994): *Environmental policy, plant location and government protection*. Springer.

ULPH, A., AND L. VALENTINI (1997): “Plant location and strategic environmental policy with inter-sectoral linkages,” *Resource and Energy Economics*, 19(4), 363–383.

VARTIAINEN, H. (2013): “Auction design without commitment,” *Journal of the European Economic Association*, 11(2), 316–342.

WEITZMAN, M. L. (1980): “The ’Ratchet Principle’ and Performance Incentives,” *Bell Journal of Economics*, 11(1), 302–308.

WILSON, J. D., AND D. E. WILDASIN (2004): “Capital tax competition: bane or boon,” *Journal of Public Economics*, 88(6), 1065–1091.