Exit Options in Incomplete Contracts with Asymmetric Information*

Helmut Bester† and Daniel Krähmer‡

June 9, 2009

Abstract

This paper analyzes bilateral contracting in an environment with contractual incompleteness and asymmetric information. One party (the seller) makes an unverifiable quality choice and the other party (the buyer) has private information about its valuation. A simple exit option contract, which allows the buyer to refuse trade, achieves the first–best in the benchmark cases where either quality is verifiable or the buyer's valuation is public information. But, when unverifiable and asymmetric information are combined, exit options induce inefficient pooling and lead to a particularly simple contract. Moreover, it is shown that simple exit options remain optimal if more general contracts are allowed under which the terms of trade are contingent on the exchange of messages between both parties.

Keywords: Incomplete Contracts, Asymmetric Information, Exit Options
JEL Classification No.: D82, D86, L15

*We wish to thank Oliver Gürtler, Paul Heidhues, David Martimort, Tymofiy Mylovanov, Roland Strausz and Klaus Schmidt for their comments and suggestions. Support by the German Science Foundation (DFG) through SFB/TR 15 is gratefully acknowledged.

†Free University Berlin, Dept. of Economics, Boltzmannstr. 20, D-14195 Berlin (Germany); email: *hbester@wiwiss.fu-berlin.de*

‡University of Bonn, Hausdorff Center for Mathematics and Institute for Theoretical Economis, Adenauer Allee 24-42, D-53113 Bonn (Germany); email: *kraehmer@hcm-uni.bonn.de*

1 Introduction

This paper analyzes bilateral contracting in environments with two potential contracting imperfections: one party has to take a decision which, though observable, is publicly not verifiable, and the other party receives decision relevant private information. The environment is thus characterized by contractual incompleteness *and* asymmetric information.[1] The parties' contracting problem is to provide incentives both for the informed party to reveal its private information and for the other party not to abuse its discretion that arises due to the lack of verifiability.

The existing literature provides core insights on what contracting can achieve if only one of the two imperfections, either non–verifiability or asymmetric information, prevails. The literature on implementation under complete information (Maskin (1977), Moore and Repullo (1988)) has studied the extent to which contracting can overcome problems caused by non-verifiable information, while the Revelation Principle (Myerson (1979)) represents the key tool to describe the set of implementable outcomes in the presence of asymmetric information. Yet little is known about how contracting is affected by the combination of unverifiable and asymmetric information. This paper presents a step in this direction.

We consider a model with a seller who has to make a non–verifiable quality choice and a buyer whose valuation for quality is his private information. There is a continuum of buyer types and the efficient level of quality is a strictly increasing function of the buyer's type. Quality is publicly not verifiable (neither ex ante nor ex post), but we assume that it is observable by the buyer. Consequently, quality cannot be legally enforced and so the seller has only imperfect commitment.

To focus on the interaction between non–verifiability and asymmetric information, we consider an environment in which the buyer learns his information only after contracting has been completed. This implies that first–best efficiency can be attained in either of the two benchmark cases in which merely one of the imperfections is present. Indeed, in the benchmark cases, first–best efficiency can be attained by an exit option contract which gives the buyer the right, after having observed the seller's quality choice, to refuse or accept to trade at a pre–specified price.[2]

[1] To fix terminology, we refer to an action as non–contractible or non–verifiable, if it is observable by the contracting parties but not verifiable to outsiders, in particular the courts. We will not consider moral hazard or hidden actions, which can be observed only by the party who controls them.

[2] It is well–known from the incomplete contracts literature that contracts with pre–specified default options can resolve obstacles that arise from non–verifiability. See, e.g., Chung (1991), Aghion, Dewatripont and Rey

Our first insight is that exit option contracts achieve no longer the first–best when non–verifiability and asymmetric information are combined. In fact, we demonstrate that exit options can implement at most a single positive level of quality and can sort buyer types in at most two groups: low valuation types will not trade the good, and high buyer types will trade the same quality of the good. Thus, while first–best efficiency calls for a perfect sorting of types, pooling of buyers is unavoidable under exit option contracts.

In a further step, we characterize the optimal exit option contract. Since only a single quality level can be implemented under an exit option contract, the optimal contract can be derived from a straightforward maximization problem, which represents a substantial simplification of the seller's original mechanism design problem.

In the final step of our analysis we provide a justification for the use of exit option contracts. Such contracts are somewhat restrictive because they limit the communication between the parties to a 'trade'– or 'exit'–message by the buyer after he observes the seller's choice of quality. We remove this restriction by considering contracts that allow conditioning the trading outcome on arbitrary forms of verifiable communication, which takes place after the buyer has observed the seller's quality choice. We demonstrate that the optimal contract has the form of an exit option when we apply the concept of strong implementation, which in the case of multiple equilibria requires all Nash equilibria of the message game to induce the same trading outcome.

To understand why lack of verifiability and asymmetric information prevent achieving first–best efficiency, it is useful to understand why efficient exit options can be designed in our two benchmark cases. If the buyer's valuation is public information, the optimal exit option leaves the buyer indifferent between exit and trade at the efficient quality level. This induces the seller to choose the efficient quality since a downward deviation would trigger the buyer to exit. In contrast, when information is private and the seller can commit to quality, the standard revelation principle is applicable, and a contract specifies a quality contingent on (a report about) the buyer's type. Incentive compatibility then requires that higher buyer types obtain a higher utility *ex post* which implies that higher buyer types must strictly prefer trade over exit for otherwise low types could gain by claiming to be a high type and then simply exiting.

Therefore, there is a tension between providing first–best incentives jointly for the seller

(1994), Nöldeke and Schmidt (1995), Edlin and Reichelstein (1996), Evans (2008). Option contracts are frequently observed in practice. For example, almost all labor contracts give the employee the right to quit. Also, certain financial contracts such as convertible bond securities can be interpreted as exit option contracts.

and the buyer. While limited commitment by the seller requires all buyer types to be indifferent between trade and exit, incentive compatibility requires (almost all) buyer types to prefer trade over exit. Thus, the constraints that arise from limited commitment and private information cannot be met jointly by an exit option contract without violating efficiency.

To characterize the set of feasible exit option contracts under asymmetric information, we allow the buyer to provide information about the realization of his valuation. After having privately observed his type, the buyer sends a verifiable message to the seller who then selects a quality level. Since quality is non–verifiable, we cannot appeal to the standard Revelation Principle and, instead, allow for general, not only direct, communication. Two forces drive the fact that not more than a single positive quality level can be implemented. First, refusing to trade has the same value for any buyer type. Second, the seller's limited commitment implies that for any positive quality level that is implemented in equilibrium, there must be some type who is indifferent between refusing and accepting trade at this quality level. Thus, if two positive quality levels are implemented, the lower of the two indifferent types could attain the same utility as the higher one by announcing the respective message and then exit. But this would contradict the incentive compatibility requirement that lower types get a lower utility than higher types in equilibrium.

A similar force underlies our result that exit options cannot improve upon more general forms of contracts that condition on communication by both parties. To induce the seller to choose a given quality, the message game following the choice of quality must deter the seller from lowering his quality. This means that after any arbitrarily small deviation from the given quality, the message game must have an equilibrium which makes the seller worse off than when he does not deviate. This essentially implies that if the seller has chosen the desired quality the buyer is indifferent between his equilibrium message and the message that he would play, did the seller slightly deviate from the desired quality. This indifference condition is akin to the indifference condition described in the previous paragraph. Together with incentive compatibility it implies that at most a single positive quality level can be implemented even when the contract conditions on communication between both parties. But a single positive quality can as well be implemented by an exit option contract.

Related Literature

This paper contributes to the literature by combining implementation under complete and incomplete information, which the existing literature largely treats as separate domains. The basic idea of implementation under complete information is that the information that the parties commonly observe can be reflected in verifiable messages to a third party.[3] A contract may therefore specify an outcome as a function of such messages and thus provide appropriate incentives for parties to select non–verifiable actions ex ante. Indeed, the efficient exit option mechanism of our first benchmark case in which the buyer's valuation is public information is an example of a sequential mechanism in the spirit of subgame perfect implementation (cf. Che and Hausch (1999), Proposition 1). However, in an environment in which there is not only non–verifiable but also asymmetric information at the communication stage, we cannot apply implementation results that rely on complete information. Instead, we study which trading outcomes can be implemented as a Bayesian Nash equilibrium after the seller has chosen quality. In the spirit of Maskin (1977), we require strong implementation and demonstrate that the combination of private and unverifiable information severely restricts the range of implementable outcomes. Importantly, since we assume contracting to take place under symmetric information, the first–best can be achieved in our other benchmark case in which quality is verifiable. Therefore, rather than the buyer's power to extract information rents, it is the lack of verifiability in combination with asymmetric information that generates inefficiencies.

Reversely, the predominant focus of the literature on implementation under incomplete information has been how to elicit private information when contracts are complete. The standard Revelation Principle (see e.g. Myerson (1979)) states that the range of implementable outcomes coincides with the set of outcomes that can be achieved through direct and truthful communication. Yet since our model displays contractual incompleteness, we cannot rely on this principle because it requires the contracting parties to write a complete contract in the sense that all message–dependent variables are specified as part of the mechanism.[4] As Bester and Strausz (2001, 2007) show, if this requirement is not satisfied, the optimal mechanism may use some form of noisy communication with only partial information revelation. Indeed, for our analysis of optimal exit options we can apply the framework of Bester and Strausz (2001), except for the technical problem that we do not consider a finite type space. In our context, partial information revelation actually simplifies the optimal contract because it pools the continuum of buyer types into merely two groups: all types below a critical type do not trade, and all other types purchase the same quality.

Our work is also related to the large literature on the hold–up problem. The key difference is that in line with much of the literature on implementation, we assume that the parties can

[3]See the seminal papers by Maskin (1977) and Moore and Repullo (1988). For a survey, see Moore (1992).

[4]In our model, this would require that the seller's quality choice is contractually determined as a function of the buyer's report about his valuation.

commit not to renegotiate ex post inefficient outcomes.[5] In contrast, the hold–up literature has studied what contracts can achieve in the absence of this commitment. Our setup can be seen as a hold–up problem where the seller's quality choice corresponds to a 'purely cooperative' ex ante investment that enhances the buyer's valuation, and the buyer does not invest. In the context of an exit option contract, our commitment assumption means that the parties can commit not to renegotiate the pre-specified terms of trade if the buyer exerts the exit option while gains from trade would exist.

While some authors argue that contract renegotiation leads to inefficient investments by substantially or even fully undermining the power of contracting (Hart and Moore (1988), Che and Hausch (1999), Edlin and Hermalin (2007)), others have identified contractual devices that induce first–best investments (Chung (1991), Aghion, Dewatripont and Rey (1994), Nöldeke and Schmidt (1995), Edlin and Reichelstein (1996), Evans (2006, 2008)). Our paper is complementary to this debate. It provides an inefficiency result which is not rooted in the parties' lack of commitment to enforce ex post inefficient default outcomes. Since the inefficiencies associated with unverifiable investments are important for providing explanations for different economic institutions (e.g. Grossman and Hart (1986), Hart and Moore (1990)), our analysis suggests that enriching the incomplete contracts paradigm by the consideration of asymmetric information may be a fruitful direction for the analysis of organizations.

Finally, our paper is related to Schmitz (2002) who establishes an inefficiency result in a bilateral trade model when the seller chooses some unobservable ex ante investment which affects the buyer's private valuation of the good.[6] The main difference is that in Schmitz' model the seller's action is not observable by the buyer, that is, there is moral hazard, whereas in our case, the seller's action is not verifiable but observable. Thus, in Schmitz' environment there is no scope for eliciting the seller's action through messages by the both parties.

This paper is organized as follows. Section 2 describes the contracting environment. In Section 3 we consider exit option contracts in the benchmark cases, where either quality is verifiable or the buyer's valuation is public information. Section 4 studies the optimal exit option contract with private *and* unverifiable information. Section 5 extends the analysis by considering messages games. Section 6 provides concluding remarks. The proofs of all formal results are relegated to an appendix in Section 7.

[5] For implementation and renegotiation under complete information see Maskin and Moore (1999).

[6] Hori (2006) and Zhao (2008) confirm Schmitz' result in mathematically more general environments.

2 The Model

We consider a buyer and a seller, who are both risk neutral. In the first stage $t = 0$ they can write a contract about the terms of trade, which occurs in some future stage $t = 3$. After a contract has been signed, the realization of a random variable θ determines the buyer's type in stage $t = 1$. In stage $t = 2$ the seller selects the quality $q \geq 0$ of an indivisible good. The buyer's valuation of consuming quality q depends on his type θ and is given by $v(q, \theta)$. The seller's cost of producing quality q is $c(q)$. In stage $t = 3$ the buyer observes the seller's quality choice. Figure 1 summarizes the sequence of events.

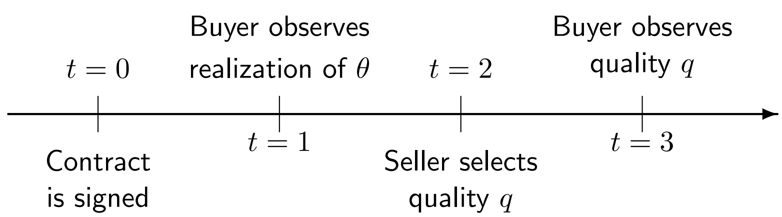

Figure 1: The Sequence of Events

In the first step of the analysis we study what the parties can achieve by using exit option contracts.[7] An exit option contract allows the buyer in stage $t = 3$ to decide whether to accept delivery or to reject and exit. We assume that the buyer's decision is publicly observable. Thus at $t = 0$ it is possible to write a contract that specifies the buyer's payment $p = (p_T, p_N)$ contingent on whether trade takes place or not.[8] Note that we do not rule out payments from the seller to the buyer because p_T and p_N are not restricted to be non–negative. Also note that the buyer's exit option in stage $t = 3$ is endogenously determined by the contract. A contract can eliminate this option simply by specifying a sufficiently large payment p_N.[9]

The buyer's (gross) outside option value is zero, independently of his type θ. Therefore, type θ accepts trade as long as $v(q, \theta) - p_T \geq -p_N$. We denote the buyer's decision behavior

[7] In Section 5 we extend the analysis and allow for contracts that condition on communication between both parties.

[8] In principle, a contract could also require the buyer to make some down–payment p_0 in stage $t = 0$. But, it is easy to see that this would be equivalent to setting $p'_T = p_T + p_0$ and $p'_N = p_N + p_0$.

[9] In contrast, Compte and Jehiel (2007) define *quitting rights* by requiring that transfers are zero in the disagreement case.

in the final stage by

$$h(q,p\,|\,\theta) = \begin{cases} 1 & \text{if} \quad v(q,\theta) - p_T \geq -p_N, \\ 0 & \text{if} \quad v(q,\theta) - p_T < -p_N. \end{cases} \tag{1}$$

Thus, the buyer type θ's payoff depends on q and p according to

$$\begin{aligned} U(q,p\,|\,\theta) &= h(q,p\,|\,\theta)[v(q,\theta) - p_T] - (1 - h(q,p\,|\,\theta))p_N \\ &= \max[v(q,\theta) - p_T, -p_N], \end{aligned} \tag{2}$$

The seller's profit is

$$\Pi(q,p\,|\,\theta) \;=\; h(q,p\,|\,\theta)p_T + (1 - h(q,p\,|\,\theta))p_N - c(q) \tag{3}$$

when he faces a buyer of type θ.

The buyer's type θ is drawn from the interval $\Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$ according to the continuously differentiable cumulative distribution function $F(\cdot)$ with $F'(\theta) > 0$ for all $\theta \in \Theta$. Let $\mathcal{T}$ denote the Borel σ–algebra on Θ. We make the following assumptions about $v(\cdot)$ and $c(\cdot)$:[10]

$$v(0,\theta) = 0,\; v_q(q,\theta) > 0,\; v_\theta(q,\theta) > 0,\; v_{qq}(q,\theta) \leq 0,\; v_{q\theta}(q,\theta) > 0, \tag{4}$$

$$c(0) = 0,\; c'(q) > 0,\; c''(q) > 0. \tag{5}$$

Finally, to avoid corner solutions, we assume that $v_q(0,\theta) > c'(0)$ and $v_q(\bar{q},\theta) < c'(\bar{q})$ for $\bar{q}$ sufficiently large.

Our assumptions ensure that for any realization of $\theta \in \Theta$ the first–best quality, which maximizes the joint surplus,

$$\tilde{q}(\theta) \equiv \operatorname{argmax}_{q \geq 0} v(q,\theta) - c(q) \tag{6}$$

is positive and unique. Also, by the last condition in (4), $\tilde{q}(\cdot)$ is strictly increasing in θ. If, in addition to the transfers p, the buyer and the seller were able to contractually specify the quality-level $\tilde{q}(\theta)$ contingent upon the realization of θ, this would maximize their ex ante expected total surplus in stage $t = 0$.

In what follows, however, we consider two limitations on the parties' contracting possibilities that prevent them from making $\tilde{q}(\theta)$ part of the contract. First, we assume that,

[10]Subscripts are used to denote partial derivatives.

although quality q is perfectly observable by both parties, it is not verifiable to outsiders. Thus a contract that explicitly specifies some q cannot be enforced by the courts. The buyer and the seller can only write an *incomplete contract* that leaves the selection of q at the seller's discretion.

Second, we assume that the buyer is privately informed about his type θ. This problem of *asymmetric information* makes it impossible to condition the variables of the contract directly upon the buyer's observation of θ. But, a contract may specify a set M of verifiable messages and require the buyer to select a message $m \in M$ after observing his type. An exit option contract (M, p) thus consists of a message set M and message contingent transfers $p: M \to \mathbb{R}^2$ such that, when in stage $t = 1$ the buyer reports $m \in M$, he has to pay $p_T(m)$ in stage $t = 3$ if accepting trade and $p_N(m)$ otherwise. Upon receiving the message m, the seller updates his beliefs about the buyer's type and chooses some quality $q(m)$ in stage $t = 2$.

The objective of our analysis is to characterize the contract that maximizes the seller's expected profit in $t = 0$ subject to the buyer's participation constraint and the restrictions imposed by contractual incompleteness and asymmetric information. But we relegate the derivation of the optimal exit option contract to Section 4. In the following section, we first consider two benchmark environments where either the quality q is contractible *or* the buyer's type θ is publicly observable.

3 Two Benchmarks

To disentangle the implications of contractual incompleteness and asymmetric information, we consider two reference points in this section. We first derive the seller's optimal contract when quality is verifiable and contractible, but the buyer's type is private information. We then analyse the case where the buyer's type is publicly observable, but quality is not verifiable. It will turn out that in either situation the seller can appropriate the first–best surplus

$$\tilde{S} \equiv \int_{\underline{\theta}}^{\bar{\theta}} [v(\tilde{q}(\theta), \theta) - c(\tilde{q}(\theta))] \, dF \tag{7}$$

under a contract that induces the buyer of type θ to accept quality $\tilde{q}(\theta)$.

Contractible q, asymmetric information about θ

Suppose quality q is verifiable so that the seller can contractually commit to $q(m)$ after receiving the buyer's message $m \in M$. In this situation, the Revelation Principle (see e.g.

Myerson (1979)) allows restricting the analysis to direct and truthful communication. Therefore, without loss of generality, the seller can use a contract with $M = \Theta$, $q\colon \Theta \rightarrow \mathbb{R}_+$ and $p\colon \Theta \rightarrow \mathbb{R}^2$. Further, the contract has to be incentive–compatible so that reporting truthfully is optimal for each type θ of the buyer.

The seller's problem is thus to maximize his expected profit subject to the incentive–compatibility conditions and the buyer's participation constraint:

$$\max_{\{q(\cdot),\, p(\cdot)\}} \int_{\underline{\theta}}^{\bar{\theta}} \Pi(q(\theta), p(\theta) \,|\, \theta) dF \tag{8}$$

subject to

$$U(q(\theta), p(\theta) \,|\, \theta) \geq U(q(\theta'), p(\theta') \,|\, \theta) \quad \text{for all } (\theta, \theta'), \tag{9}$$

$$\int_{\underline{\theta}}^{\bar{\theta}} U(q(\theta), p(\theta) \,|\, \theta) dF \geq 0. \tag{10}$$

The incentive compatibility constraints (9) ensure that no buyer has an incentive to misrepresent his type. Note that our incentive compatibility constraints are somewhat non–standard, compared e.g. to a standard price discrimination problem, because they also comprise that no buyer has an incentive to misreport his type and subsequently refuse to trade. The participation constraint (10) guarantees that the buyer's expected utility at the contracting stage, before he learns his type, is at least zero. The next proposition states that the first–best can be implemented.

Proposition 1 *(a) There exists a $p^*(\cdot)$ such that $\{\tilde{q}(\cdot), p^*(\cdot)\}$ solves problem (8) – (10). Moreover, $h(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) = 1$ for all $\theta \in \Theta$ and*

$$\int_{\underline{\theta}}^{\bar{\theta}} \Pi(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) dF = \tilde{S}.$$

(b) For any solution $\{q^*(\cdot), p^*(\cdot)\}$ of problem (8) – (10) it holds for almost all $\theta \in \Theta$ that

$$v(q^*(\theta), \theta) - p^*_T(\theta) > -p^*_N(\theta).$$

The idea behind part (a) is to specify a large exit payment so that no buyer type wants to exit. This effectively eliminates the exit option and we are back in a standard price discrimination framework for which it is well known that the seller can fully extract the first–best surplus if the buyer learns his private information only ex post.

Part (b) is an implication of incentive compatibility for the buyer's trade incentives that any optimal contract has to satisfy. In light of (a), any optimal contract must extract all gains from trade and thus induce almost all buyer types to trade. Now if two buyer types trade, a straightforward implication of incentive compatibility is that the high valuation buyer must obtain a larger ex post utility $v - p_T$ than the low valuation buyer. It follows that almost any buyer type (except possibly the lowest) must strictly prefer trade over exit after reporting his type truthfully. Otherwise, if one buyer type θ was exactly indifferent, all smaller types $\theta' < \theta$ would be better off by pretending to be type θ in $t = 1$ and exiting in $t = 3$.

Non–contractible q, public information about θ

Suppose now that the buyer's type θ is public information and that quality q, though observable by both parties, is not contractible. In this situation, messages from the buyer about his type are redundant, and the seller can simply offer a contract $p : \Theta \to \mathbb{R}^2$ where the trade and exit transfers are $p(\theta)$, when the buyer's type is θ. Since q is not contractible, the seller will select q ex post so as to maximize his profits given the transfers $p(\theta)$. In other words, the choice of q is constrained by imperfect commitment on part of the seller.

The seller's problem is thus to maximize his expected profit subject to his *no–commitment constraint* and the buyer's participation constraint:

$$\max_{\{q(\cdot),\, p(\cdot)\}} \int_{\underline{\theta}}^{\bar{\theta}} \Pi(q(\theta), p(\theta) \,|\, \theta) dF \tag{11}$$

subject to

$$\Pi(q(\theta), p(\theta) \,|\, \theta) \geq \Pi(q', p(\theta) \,|\, \theta) \quad \text{for all } q', \theta, \tag{12}$$

$$\int_{\underline{\theta}}^{\bar{\theta}} U(q(\theta), p(\theta) \,|\, \theta) dF \geq 0. \tag{13}$$

The no–commitment constraint (12) describes the seller's choice of quality in $t = 2$. He selects q to maximize his profit ex post, given the transfers p and the buyer's type θ. Thus, when designing the contract, the seller has to take into account his ex post incentives for selecting q. Even though quality cannot be contractually determined, the next proposition demonstrates that by the appropriate choice of exit options the seller can commit himself to choose the first–best quality $\tilde{q}$ ex post.

Proposition 2 *(a) There exists a $p^*(\cdot)$ such that $\{\tilde{q}(\cdot), p^*(\cdot)\}$ solves problem (11) – (13). Moreover, $h(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) = 1$ for all $\theta \in \Theta$ and*

$$\int_{\underline{\theta}}^{\bar{\theta}} \Pi(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) dF = \tilde{S}.$$

(b) For any solution $\{q^*(\cdot), p^*(\cdot)\}$ of problem (8) – (10) it holds for almost all $\theta \in \Theta$ that

$$v(q^*(\theta), \theta) - p_T^*(\theta) = -p_N^*(\theta).$$

The basic idea behind part (a) is to contract an exit payment of zero and to specify the trade transfer in such a way that each buyer type is exactly indifferent between trade and exit when the seller offers the first–best quality. This contract commits the seller not to shirk ex post because otherwise the buyer would exit and leave the seller with a zero payment. Part (b) illuminates the implications of the no–commitment constraint for the buyer's trade incentives. Under any optimal contract the buyer needs to be indifferent between exit and trade when offered the first–best quality. Otherwise, incentives would arise for the seller to shade quality below the first–best.

Proposition 2 (a) is closely related to an observation by Che and Hausch (1999) who show that the first–best can be implemented when the parties can commit themselves not to renegotiate the contract. They continue their analysis by establishing an inefficiency result if committing not to renegotiate the contract is impossible. In contrast, we maintain the assumption that contracts are not renegotiated. In the next section, we provide a different inefficiency result for the case where the buyer's type is private information. In this sense, our analysis is complementary to Che and Hausch (1999).

4 Exit Options

Part (b) of Propositions 1 and 2 are clearly incompatible: when the buyer's type is private information, each buyer type must strictly prefer trade over exit while when quality is non–contractible, each buyer type needs to be indifferent between trade and exit. Thus, there is a tension in providing appropriate incentives jointly for the buyer (incentive compatibility) and the seller (no–commitment). This indicates that the first–best cannot be implemented when the seller cannot contractually commit to some quality q and, at the same time, the buyer is privately informed about his type θ.

We now turn to characterizing the optimal exit option contract in this case. For this type of problem, it is well–known that it may not be optimal to use a direct communication mechanism that induces truthful revelation. Indeed, as shown in Bester and Strausz (2001), an indirect mechanism may support outcomes that cannot be replicated by a direct mechanism. Bester and Strausz (2001) also show, however, that when the set of types Θ is finite, any *incentive efficient* outcome can be replicated by an equilibrium of a direct mechanism. Their result, however, does not apply to our environment since the set Θ represents a continuum of types.[11] Therefore, we first characterize the outcomes that can be supported as a Perfect Bayesian Equilibrium under some arbitrary message set M. This allows us in a second step to derive the seller's optimal exit option contract.

Perfect Bayesian Equilibrium

Let the message set M be an arbitrary metric space and let $\mathcal{M}$ denote the Borel $\sigma-$algebra on M. The contract between the seller and the buyer specifies the transfers $p\colon M \to \mathbb{R}^2$. Thus, when the buyer reports $m \in M$, he has to pay $p_T(m)$ if accepting trade, and $p_N(m)$ if he exits in the final stage. The functions $p_N(\cdot)$ and $p_T(\cdot)$ are taken to be measurable.

We denote the θ–type buyer's *reporting strategy* by $r(\cdot|\theta) \in Q$, where Q is the set of probability measures on $\mathcal{M}$. Thus, if $r(H \,|\, \theta) > 0$ for some $H \in \mathcal{M}$, this means the message chosen by the θ–type buyer lies in H with probability $r(H \,|\, \theta)$.

After receiving message m, the seller updates his *beliefs* about the buyer's type. We denote these beliefs as $\mu(T, m)$. Thus, upon observing message m, the seller believes that the buyer's true type is in the set $T \in \mathcal{T}$ with probability $\mu(T, m)$. Given his beliefs, the seller chooses $q(m)$ to maximize his expected payoff.

To constitute a *Perfect Bayesian Equilibrium*, the functions (r, μ, q) have to satisfy three conditions: First, the seller's choice of q has to be optimal given his beliefs. This means that $q(\cdot)$ has to satisfy *the no–commitment constraint*

$$q(m) = \operatorname*{argmax}_q \int_{\underline{\theta}}^{\bar{\theta}} \Pi(q, p \,|\, \theta)\mu(\theta, m)d\theta \tag{14}$$

[11] Actually, the assumption of a continuum of types simplifies our analysis. In a different context also Krishna and Morgan (2008) consider a contracting problem with imperfect commitment and a continuum of types.

for all $m \in M$.

Second, as the buyer anticipates that message m will induce the seller to select $q(m)$, he will select an optimal reporting strategy. The *set of optimal messages* for type θ is

$$M(\theta) \equiv \{m \in M \,|\, U(q(m), p(m) \,|\, \theta) \geq U(q(m'), p(m') \,|\, \theta) \text{ for all } m' \in M\}. \tag{15}$$

Let $R(\theta)$ denote the *support* of the θ-type buyer's reporting strategy $r(\cdot|\theta)$. Then optimality of the buyer's reporting strategy requires that

$$R(\theta) \subseteq M(\theta) \text{ for all } \theta \in \Theta. \tag{16}$$

We refer to the constraint (16) as the buyer's *communication incentive constraint*.

Third, the seller's belief μ has to be consistent with Bayesian updating on the support of the buyer's reporting strategy. This means that $\mu(\cdot, m)$ is derived from Bayes' rule whenever $m \in R(\theta)$ for some $\theta \in \Theta$. Of course, the belief μ determines the seller's choice of q also for messages that lie outside the support of the buyer's reporting strategy. Yet, there are no consistency restrictions on beliefs for such messages.

Feasible contracts

Our next aim is to characterize the equilibrium outcomes that can arise under an arbitrary contract (M, p). We demonstrate that at most a single positive quality level can be implemented in equilibrium. Let us begin by introducing further notation. Consider a Perfect Bayesian Equilibrium under some arbitrary message set M. In equilibrium, each buyer type submits a message m and will then be offered the quality $q(m)$. We say that *trade at a positive quality* takes place if $q(m) > 0$ and the buyer accepts to trade. We denote by $M^+(\theta) \subseteq M(\theta)$ the set of all messages that are optimal for the θ–type buyer and lead to trade at a positive quality:

$$M^+(\theta) \equiv \{m \in M(\theta) \,|\, q(m) > 0 \text{ and } h(q(m), p(m) \,|\, \theta) = 1\}. \tag{17}$$

We denote by $R^+(\theta) \subseteq R(\theta)$ the set of all messages that are in the support of the θ–type buyer and lead to trade at a positive quality:

$$R^+(\theta) \equiv R(\theta) \cap M^+(\theta). \tag{18}$$

If $m \in R^+(\theta)$, we refer to m as a *positive trade message for buyer type θ*. For a given message m, we collect all types for whom m is a positive trade message in the set $T^+(m)$:

$$T^+(m) \equiv \{\theta \in \Theta \,|\, m \in R^+(\theta)\}. \tag{19}$$

Notice that $T^+(m) = \emptyset$ if and only if there is no buyer type for whom m is a positive trade message, that is, m is in no buyer type's support, or $q(m) = 0$, or each buyer who submits m exits. Therefore, we refer to m as a *positive trade message* if $T^+(m) \neq \emptyset$. For any positive trade message, we define

$$\theta_\ell(m) \equiv \inf T^+(m). \tag{20}$$

The next two lemmas state basic consequences of the no–commitment (14) and the communication incentive (16) constraints. Lemma 1 follows from (14).

Lemma 1 *Let m be a positive trade message, then the buyer type $\theta_\ell(m)$ is indifferent between trade and exit, i.e. $v(q(m), \theta_\ell(m)) - p_T(m) = -p_N(m)$ if $T^+(m) \neq \emptyset$.*

To see the intuition for Lemma 1, note that each type for whom m is a positive trade message, weakly prefers trade over exit conditional on reporting m. Thus, by continuity, also the type $\theta_\ell(m)$ weakly prefers trade over exit when offered $q(m)$. The fact that he cannot strictly prefer trade over exit is a consequence of the seller's no–commitment constraint: when receiving message m, the seller infers that the buyer's type cannot be smaller than $\theta_\ell(m)$ because no type smaller than $\theta_\ell(m)$ sends message m in equilibrium. Thus, if the $\theta_\ell(m)$–type strictly preferred trade over exit, the seller could slightly reduce the quality and the buyer would still accept to trade with probability 1.

The next lemma follows from Lemma 1 and the communication incentive constraint.

Lemma 2 *The exit payments $p_N(m)$ and the types $\theta_\ell(m)$ are the same for all positive trade messages m, i.e. $p_N(m) = p_N(m')$ and $\theta_\ell(m) = \theta_\ell(m')$ if $T^+(m) \neq \emptyset$ and $T^+(m') \neq \emptyset$.*

To understand Lemma 2, observe first that continuity of U in θ and the definition of the infimum imply that any positive trade message m is an optimal message for the buyer type $\theta_\ell(m)$. Since $\theta_\ell(m)$ is indifferent between exit and trade when he sends message m, his utility from sending m is simply $-p_N(m)$. Hence, if there was some other message m' with $p_N(m') < p_N(m)$, message m could not be optimal, as submitting m' and exiting would yield the buyer a larger utility.

Further, the intuition for why $\theta_\ell(m) = \theta_\ell(m')$ is similar to the case in which q is contractible. If two buyer types weakly prefer trade over exit upon sending some message, then the higher type must obtain a strictly larger utility $v - p_T$ in order for him not to have incentives to deviate to the message of the lower type. Hence, $\theta_\ell(m)$ must be the same as

$\theta_\ell(m')$ because by Lemma 1 both types weakly prefer to trade and their utility $v - p_T$ is the same due to Lemma 1 and because $p_N(m') = p_N(m)$.

Lemma 2 allows us to define a *critical type* and constant exit payments for all positive trade messages m:[12]

$$\hat{\theta} \equiv \theta_\ell(m) \text{ and } \hat{p}_N \equiv p_N(m) \text{ for all } m \text{ with } T^+(m) \neq \emptyset. \tag{21}$$

From Lemmas 1 and 2 we deduce:

$$v(q(m), \hat{\theta}) - p_T(m) = -\hat{p}_N \text{ for all } m \text{ with } T^+(m) \neq \emptyset. \tag{22}$$

Condition (22) says that only such positive quality levels can be implemented as an equilibrium for which the critical type is indifferent between trade and exit. In fact, the no–commitment and communication incentive constraints together imply that only a single positive quality level can be implemented in equilibrium. This is stated in the following equilibrium characterization:

Proposition 3 *In any Perfect Bayesian Equilibrium, there is a $\hat{\theta}$ and a $\hat{q} > 0$ such that:*

(i) For all $\theta > \hat{\theta}$ and $m \in R(\theta)$ it holds that $q(m) = \hat{q}$ and $h(q(m), p(m) \,|\, \theta) = 1$.

(ii) For all $\theta < \hat{\theta}$ and $m \in R(\theta)$ it holds that $q(m) = 0$ or $h(q(m), p(m) \,|\, \theta) = 0$.

The proposition says that in equilibrium only an imperfect sorting of types into two groups can occur and that at most one group can trade at a positive quality level. A finer sorting of types, say with two positive quality levels, is impossible because communication incentives would imply that the high quality traders must get a higher utility from trade than the low quality traders. At the same time, for high quality provision by the seller to be credible, the lowest high quality trader must be indifferent between trade and exit. But then a low quality trader can obtain the same utility as this high quality trader by asking for the high quality and then exiting.

An immediate corollary of Proposition 3 is that the first–best cannot be implemented. By Propositions 1 and 2, this inefficiency result is driven by the combined presence of private information and contractual incompleteness.

[12] If there is no positive trade message, i.e. if $T^+(m) = \emptyset$ for all $m \in M$, we set $\hat{\theta} = \bar{\theta}$.

Optimal Exit Options

We now derive the optimal exit option contract for the seller. Proposition 3 implies that the optimal contract can be found in the class of contracts that have only two messages, say m_l, m_h. Such a contract induces a Perfect Bayesian Equilibrium in which all 'high' types above a critical $\hat{\theta}$ report the message m_h and trade the positive quality $q(m_h) = \hat{q}$, and all 'low' types below $\hat{\theta}$ report message m_l and trade a zero quality $q(m_l) = 0$.

The seller's problem is to choose transfers $p = \{p_N(m_l), p_T(m_l), p_N(m_h), p_T(m_h)\}$, a quality $\hat{q}$, and a critical type $\hat{\theta}$ that maximize his ex ante profit subject to the participation constraint and the constraint that $(\hat{q}, \hat{\theta})$ can be supported as a Perfect Bayesian Equilibrium given the transfers p. Without loss of generality, we set $p_N(m_l) = p_T(m_l) = p_N(m_h)$ and define $p_N = p_N(m_h)$, and $p_T = p_T(m_h)$ with $p_T > p_N$.[13] Formally, the seller's problem is:

$$\max_{p_N, p_T, \hat{q}, \hat{\theta}} F(\hat{\theta})p_N + (1 - F(\hat{\theta}))(p_T - c(\hat{q})) \tag{23}$$

subject to

$$\hat{q} \in \operatorname*{argmax}_{q} \int_{\hat{\theta}}^{\bar{\theta}} \frac{\Pi(q, p \mid \theta)}{1 - F(\hat{\theta})} \, dF(\theta), \tag{24}$$

$$v(\hat{q}, \hat{\theta}) = p_T - p_N, \tag{25}$$

$$-F(\hat{\theta})p_N + \int_{\hat{\theta}}^{\bar{\theta}} [v(\hat{q}, \theta) - p_T] \, dF(\theta) \geq 0. \tag{26}$$

The seller's objective (23) consists of two parts. The first part is the expected profit that he extracts from the types who announce message m_l and pay the transfer p_N. Since the quality traded is zero, no production costs accrue to the seller in this case. The second part is the expected profit that the seller extracts from the types who announce message m_h and pay the transfer p_T. Since all these types trade quality $\hat{q}$, the seller has costs $c(\hat{q})$ in this case.

Constraints (24) and (25) require that $(\hat{q}, \hat{\theta})$ constitutes an equilibrium. By the no–commitment constraint (24), if the seller receives message m_h, he infers that the buyer type is larger than $\hat{\theta}$, and his belief that he faces a type θ is given by the conditional distribution $dF(\theta)/(1 - F(\hat{\theta}))$. Given these beliefs, $\hat{q}$ has to be the optimal quality selection. Condition

[13] By Proposition 3, all types who send message m_l trade quality 0 and exit. Hence, we need that $v(0, \theta) - p_T(m_l) \leq -p_N(m_l)$, and that the seller optimally set $q = 0$ if he receives message m_l. Any transfers with $p_N(m_l) = p_T(m_l)$ satisfy these two requirements. Further, equating $p_T(m_l)$ and $p_N(m_h)$ is a normalization. Finally, $p_T > p_N$ because by Proposition 3, we must have: $v(\hat{q}, \hat{\theta}) - p_T = -p_N$. Since $\hat{q} > 0$, this implies that $p_T > p_N$.

(25) is the equilibrium requirement from Proposition 3 that the critical type $\hat{\theta}$ be indifferent between exit and trade at transfers p and quality level $\hat{q}$. Finally, (26) is the buyer's ex ante participation constraint.

We proceed by making the seller's problem more tractable. Observe first that the participation constraint must obviously be binding at the optimum. Combining this with the seller's objective, the seller's problem becomes to maximize the total surplus

$$S(\hat{q},\hat{\theta}) = \int_{\hat{\theta}}^{\bar{\theta}} (v(\hat{q},\theta) - c(\hat{q}))\, dF(\theta) \tag{27}$$

subject to (24) and (25).

Next, we reformulate the constraints (24) and (25). As explained above, these constraints embody the two requirements that the seller's choice be optimal given his beliefs, and that the seller's beliefs be consistent with the buyer's reporting strategy. To describe equilibrium, we first consider the seller's optimal quality choice (his 'best response') against arbitrary beliefs. Suppose the seller has received message m_h and holds the belief that all types larger than an *arbitrary* type $\hat{\theta}$ have submitted m_h. Then choosing a relatively high quality q with $v(q,\hat{\theta}) - p_T > -p_N$ is clearly suboptimal for him, because all types $\theta \geq \hat{\theta}$ have a strict incentive to trade and so the seller could gain by slightly lowering quality. Therefore, the seller must optimally choose a quality level q' such that $v(q',\hat{\theta}) - p_T \leq -p_N$. By setting such a quality q', the seller effectively chooses a type $\theta' \in [\hat{\theta},\bar{\theta}]$ who is indifferent between trade and exit because $v(q',\theta') = p_T - p_N$. All types $\theta \geq \theta'$ accept quality q', whereas all types $\theta \in [\hat{\theta},\theta']$ exit. Thus, the seller anticipates that quality q' will be rejected with probability $(F(\theta') - F(\hat{\theta}))/(1 - F(\hat{\theta}))$ and accepted with probability $(1 - F(\theta'))/(1 - F(\hat{\theta}))$.

Thus, given transfers p and given the belief that all types larger than type $\hat{\theta}$ have submitted m_h, the seller's optimal behavior is defined by the pair

$$(q^*(\hat{\theta},p), \theta^*(\hat{\theta},p)) \qquad \equiv \operatorname*{argmax}_{q',\theta'} \frac{F(\theta') - F(\hat{\theta})}{1 - F(\hat{\theta})} p_N + \frac{1 - F(\theta')}{1 - F(\hat{\theta})} p_T - c(q') \tag{28}$$

$$\text{subject to} \quad v(q',\theta') = p_T - p_N \quad \text{and} \quad \theta' \geq \hat{\theta}. \tag{29}$$

While (28) describes the seller's best response against arbitrary beliefs, in equilibrium the seller's beliefs are consistent with the buyer's actual behavior. This is made explicit in the next lemma which provides an alternative characterization of the equilibrium conditions (24) and (25).

Lemma 3 *Let p be given. Then $(\hat{q}, \hat{\theta})$ satisfies (24) and (25) if and only if $(\hat{q}, \hat{\theta})$ solves the following fixed–point problem:*

$$\hat{q} = q^*(\hat{\theta}, p) \quad \text{and} \quad \hat{\theta} = \theta^*(\hat{\theta}, p). \tag{30}$$

Since the conditions (30) include the optimality conditions for the seller, $(\hat{q}, \hat{\theta})$ has to satisfy the necessary first–order conditions for optimality of problem (28). We now impose conditions on F and v such that the first–order conditions are actually sufficient for optimality. This allows us to state Lemma 3 in terms of first–order conditions.

Lemma 4 *Let $F(\cdot)$ be convex and $v(\cdot)$ be quasi–concave.*[14] *For given p, $(\hat{q}, \hat{\theta})$ then solves the fixed–point problem (30) if and only if*

$$-\frac{F'(\hat{\theta})}{1 - F(\hat{\theta})}(p_T - p_N) + c'(\hat{q})\frac{v_\theta(\hat{q}, \hat{\theta})}{v_q(\hat{q}, \hat{\theta})} \leq 0, \tag{31}$$

$$v(\hat{q}, \hat{\theta}) = p_T - p_N. \tag{32}$$

Finally, we can eliminate transfers from the seller's problem. To see this, note that for any $(\hat{q}, \hat{\theta})$, transfers can be found such that (32) holds. Hence, we can insert (32) in (31) and obtain a single constraint that is independent of transfers. Since the objective $S(\hat{q}, \hat{\theta})$ is also independent of transfers, the seller's problem reduces to a maximization problem just over $(\hat{q}, \hat{\theta})$. The next proposition summarizes our findings.

Proposition 4 *Let $F(\cdot)$ be convex and $v(\cdot)$ be quasi–concave. Then the seller's problem is*

$$\max_{\hat{q}, \hat{\theta}} \int_{\hat{\theta}}^{\bar{\theta}} [v(\hat{q}, \theta) - c(\hat{q})]\, dF(\theta) \tag{33}$$

subject to

$$-\frac{F'(\hat{\theta})}{1 - F(\hat{\theta})} v(\hat{q}, \hat{\theta}) + c'(\hat{q})\frac{v_\theta(\hat{q}, \hat{\theta})}{v_q(\hat{q}, \hat{\theta})} \leq 0. \tag{34}$$

In other words, the feasible set of $(\hat{q}, \hat{\theta})$–combinations which jointly satisfy the seller's no–commitment and the buyer's incentive communication incentive constraints reduces to the simple inequality constraint (34).[15] This is a rather remarkable simplification of the problem that we started out with.

[14] A sufficient condition for $v(\cdot)$ to be quasi–concave is that $v_{\theta\theta} \leq 0$ in addition to the assumptions in (4).

[15] The corresponding transfers p_N and p_T are determined by (25) and (26).

Using Proposition 4, it is straightforward to compute the optimal contract. Let, for example, θ be uniformly distributed on $[0, 1]$, $v(q, \theta) = q\theta$, and $c(q) = cq^2$. Then it is easily verified that $\hat{q} = 0.3492/c$ and $\hat{\theta} = 0.5565$ solve problem (33)–(34). By (25) and (26) the optimal contract specifies the trade payment $p_T = 0.2286/c$ and the exit payment $p_N = 0.0343/c$.

5 Message Games

As our analysis in the previous section has shown, the first–best cannot be achieved by exit option contracts when the buyers' type is private information *and* quality is not verifiable. This raises the question of whether some other form of contract may be more efficient. Indeed, an exit option contract limits communication to a simple message of the buyer whether he accepts or refuses trade. Given that both parties are informed about the seller's quality choice, trade can more generally be made contingent on the outcome of a *message game* in which the parties use their information to exchange verifiable messages. Allowing for general message games considerably extends the range of possible contracts, including contracts that the implementation literature deems implausible (such as "shoot-the-liar"). We will therefore somewhat restrict the class of feasible contracts. As the main result in this section, we show that exit option contracts remain to be optimal in this class. This provides a justification for the use of exit option contracts even when trade and payments can be made contingent on verifiable reports by the seller and the buyer.

To study this more general form of contracting, we modify stage $t = 3$ of the environment described in Section 2. In $t = 3$, the seller and the buyer now become engaged in a message game, where they simultaneously select messages $z_S \in Z_S$ and $z_B \in Z_B$, respectively. Even though we describe the exchange of messages as a static game, this description may be thought of as the normal form representation of a dynamic game involving many stages of communication. The messages selected in $t = 3$ are verifiable so that the terms of trade can be contractually specified as a function of the buyer's message $m \in M$ in stage $t = 1$ and the outcome $z = (z_S, z_B) \in Z \equiv Z_S \times Z_B$ of the message game in stage $t = 3$. Thus, in addition to the message sets (M, Z), a contract specifies an *expected* payment $p(m, z)$ from the buyer to the seller. Also, it is verifiable whether trade occurs or not. This is described by the variable $x(m, z) \in \{0, 1\}$, where $x = 1$ indicates that the seller delivers the good to the buyer.[16] More formally, a contract is now a combination (M, Z, x, p), where $x \colon M \times Z \to \{0, 1\}$ and

[16]Implicitly we rule out random trade. Indeed, as argued by Che and Hausch (1999), random mechanisms are questionable to implement because of legal enforcement problems.

$p: M \times Z \to \mathbb{R}$.[17] The buyer's and the seller's expected payoffs are defined as

$$U(q, m, z|\theta) \equiv x(m, z)v(q, \theta) - p(m, z), \quad \Pi(q, m, z) \equiv p(m, z) - c(q). \tag{35}$$

It is easy to see that this environment entails the exit option contract as a special case, where x and p do not depend on the seller's message and the buyer has only two messages, with $x = 1$ for one message and $x = 0$ for the other. Note that our analysis does not allow the buyer and seller to include payments to a third party in their contract. A justification for this restriction is that three party contracts may be difficult to implement because they raise the problem of collusion between two of the agents against the third.[18]

Our description of Perfect Bayesian Equilibrium readily extends to the present context. The only novelty is that, after m and q have been selected in the previous stages, in $t = 3$ now a continuation game $\Gamma(m, q)$ starts in which the seller has imperfect information about the buyer's type. After having observed the buyer's message m, the seller enters $\Gamma(m, q)$ with the belief that the buyer's true type is in the set $T \in \mathcal{T}$ with probability $\mu(T, m)$. The game $\Gamma(m, q)$ is thus a (static) Bayesian game, and as part of the Perfect Bayesian Equilibrium of the overall game, the players' message strategies have to constitute a Bayesian Nash Equilibrium. This means that $(\hat{z}_S, \hat{z}_B(\cdot))$, with $\hat{z}_S \in Z_S$ and $\hat{z}_B: \Theta \to Z_B$, is an equilibrium of $\Gamma(m, q)$ if the seller's message $\hat{z}_S$ satisfies

$$\int_{\underline{\theta}}^{\bar{\theta}} p(m, \hat{z}_S, \hat{z}_B(\theta))\mu(\theta, m)d\theta \geq \int_{\underline{\theta}}^{\bar{\theta}} p(m, z_S, \hat{z}_B(\theta))\mu(\theta, m)d\theta \quad \text{for all } z_S \in Z_S, \tag{36}$$

and each buyer type θ with $m \in R(\theta)$ selects a message $\hat{z}_B(\theta)$ such that

$$U(q, m, \hat{z}_S, \hat{z}_B(\theta)|\theta) \geq U(q, m, \hat{z}_S, z_B|\theta) \quad \text{for all } z_B \in Z_B. \tag{37}$$

Notice that in $t = 3$ the seller's production costs are already sunk so that in (36) he only cares about expected payments when choosing his message $\hat{z}_S$. In what follows we denote by $E(m, q)$ the set of Bayesian Nash Equilibria of the game $\Gamma(m, q)$.

As is well–known, message games typically admit a multiplicity of equilibria. While some of these equilibria may implement the desired outcome, others may induce unintended outcomes. To resolve this problem, we will apply the usual concept of *strong implementation*, which requires that *all* equilibria in $E(m, q)$ have identical outcomes. In complete information

[17]In principle, Z may depend on $m \in M$. In what follows, we ignore this possibility because it does not affect our results.

[18]cf. Hart and Moore (1988), especially footnote 20.

environments the requirement of strong implementation rules out, e.g., mechanisms of the "shoot-the-liar"–type where agents are harshly punished if they announce different statements about the commonly known state. While this mechanism does have a Nash equilibrium which elicits the true state, it also has many other undesirable Nash equilibria in which agents do not tell the truth.

More specifically, we restrict the set of admissible contracts by imposing the following condition on all continuation games $\Gamma(m, q)$:

Condition 1 If $(\hat{z}_S, \hat{z}_B(\cdot)) \in E(m, q)$ and $(\tilde{z}_S, \tilde{z}_B(\cdot)) \in E(m, q)$, then

$$x(m, \hat{z}_S, \hat{z}_B(\theta)) = x(m, \tilde{z}_S, \tilde{z}_B(\theta)) \quad \text{and} \quad p(m, \hat{z}_S, \hat{z}_B(\theta)) = p(m, \tilde{z}_S, \tilde{z}_B(\theta)) \tag{38}$$

for almost all θ such that $m \in R(\theta)$.

Thus, if the buyer type θ has reported $m \in M$ in stage $t = 1$ and the seller has produced quality q in stage $t = 2$, Condition 1 implies that the trade outcome x and the payment p are uniquely determined by the outcome of the subsequent message game in $t = 3$, even when this game has multiple equilibria.

After a contract has been signed, the *path* of a Perfect Bayesian Equilibrium induces for each buyer type θ some message $m^*(\theta)$ in stage $t = 1$. Given his equilibrium beliefs $\mu^*(\cdot, m^*(\theta))$, the seller then chooses some quality $q^*(\theta)$ in $t = 2$. Finally, in $t = 3$ the equilibrium outcome $(z^*_S(\theta), z^*_B(\theta))$ of the message game $\Gamma(m^*(\theta), q^*(\theta))$ determines a probability of trade $x^*(\theta)$ and a payment $p^*(\theta)$. We say that a contract *implements* (q^*, x^*), with $q^* : \Theta \to \mathbb{R}_+$ and $x^* : \Theta \to \{0, 1\}$, if there is a Perfect Bayesian Equilibrium such that, for each type θ, equilibrium play results in production of quality $q^*(\theta)$ and trade is given by $x^*(\theta)$.[19]

For a given continuation game $\Gamma(\bar{m}, \bar{q})$ with $\bar{q} > 0$ we define the set of all buyer types who, after having sent the message $\bar{m}$ in $t = 1$, trade quality $\bar{q}$ as the outcome of the message game in stage $t = 3$:

$$\hat{T}(\bar{m}, \bar{q}) = \{\theta \in \Theta \,|\, m^*(\theta) = \bar{m}, q^*(\theta) = \bar{q} \text{ and } x^*(\theta) = 1\}. \tag{39}$$

The next lemma is key to our analysis.

[19]To simplify our analysis, we restrict ourselves to equilibria where the buyer uses a pure strategy in $t = 1$.

Lemma 5 *Suppose that Condition 1 is satisfied and that* (q^*, x^*) *can be implemented. Let* $\hat{T}(\bar{m}, \bar{q}) \neq \emptyset$ *for some* $\bar{q} > 0$*. Then for all* $\varepsilon \in (0, \bar{q})$ *there exist a* $\theta \in \hat{T}(\bar{m}, \bar{q})$ *and a message* $z'_B \neq z^*_B(\theta)$ *such that*

$$U(\bar{q} - \varepsilon, m^*(\theta), z^*_S(\theta), z^*_B(\theta)|\theta) < U(\bar{q} - \varepsilon, m^*(\theta), z^*_S(\theta), z'_B|\theta) \tag{40}$$

and $x(m^*(\theta), z^*_S(\theta), z'_B) = 0$*.*

Recall from the analysis of exit options that to implement a positive quality level, by (22) the smallest buyer type who trades this quality level has to be indifferent between trade and exit. This buyer type would therefore exit after observing that the seller has reduced quality below the equilibrium level. Lemma 5 extends this insight to the general contracting environment. The inequality in (40) states for some buyer type θ there has to be a message z'_B so that this buyer will refuse to trade if the seller deviates to a quality below $\bar{q}$. The message z'_B creates a credible threat for the buyer that deters the seller to lower quality ex post. It serves a similar function as an exit option in restraining the seller's limited commitment. In fact, as the following proposition shows, message games cannot outperform a simple exit option contract.

Proposition 5 *Suppose that Condition 1 holds and that* (q^*, x^*) *can be implemented. Then the following holds:*

(i) At most a single positive quality level can be implemented. That is, there is $\hat{\theta}$ *and a* $\hat{q} > 0$ *such that* $q^*(\theta) = \hat{q}$ *and* $x^*(\theta) = 1$ *for all* $\theta > \hat{\theta}$ *and* $q^*(\theta) = 0$ *or* $x^*(\theta) = 0$ *for all* $\theta < \hat{\theta}$*.*

(ii) (q^*, x^*) *can be implemented by an exit option contract.*

Proposition 5 shows that no efficiency gains are possible by using a more complex mechanism than a simple exit option contract. Even though we have shown in Section 2 that exit option contracts fail to achieve full efficiency, they are second–best efficient. Proposition 5 establishes a central role for exit option contracts in overcoming problems caused by contractual incompleteness and asymmetric information.

Our analysis focusses on strong implementation in Bayesian Nash equilibrium because this readily extends the concept of Nash implementation in complete information environments. It is well known that in complete information environments the use of equilibrium refinements may enrich the set of strongly implementable outcomes. As Moore and Repullo (1988) show, almost any outcome function can be implemented as a unique Subgame Perfect equilibrium of a sequential mechanism. Of course, the mechanism used by Moore and Repullo is not directly applicable to our context because it relies on complete information. But one may still ask if sequential schemes could achieve more in our setup by replacing Condition 1 for example through the alternative requirement that each continuation game have a unique Perfect Bayesian equilibrium with strategies $(\hat{z}_S, \hat{z}_B(\cdot))$ for the sequential message game. The basic argument in the proof of Lemma 5 seems to apply also under this alternative requirement. Therefore our conclusions may well go beyond the use of the Nash concept.

Note that the exit option contract described in Propositions 3 and 5 is simpler than the efficient contracts of the benchmark cases in the sense that the former supports only a single quality $\hat{q}$ rather than a schedule of type dependent qualities. Thus, as the contracting environment becomes more complex, the resulting contractual arrangement actually becomes simpler. Complex environments may therefore be consistent with the widespread use of relatively simple contracts in reality.

6 Conclusion

We have studied bilateral contracting in an environment which is characterized by both contractual incompleteness and asymmetric information. We demonstrate that the joint occurrence of these imperfections necessarily prevents exit option contracts from achieving first–best efficiency. Moreover, more general contracts cannot improve upon simple exit option contracts.

Our inefficiency result suggests that incomplete contracts with asymmetric information may be useful for studying institutional design, even in the absence of contract renegotiation. This is so because the allocation of property rights or decision rights may matter for the extent to which efficiency can be achieved. Imagine, for example, that the non–verifiable action is more broadly interpreted as some decision that an organization has to take. Suppose further that the right to take this decision can be conferred to one of its members. This assignment of authority may be enforced by the ownership of assets and resources that are necessary to implement a decision. In such an environment, the optimal institutional arrangement can be determined by comparing the efficiency implications of different allocations of property and decision rights.

7 Appendix

Proof of Proposition 1: (a) Let

$$k \equiv \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} v_\theta(\tilde{q}(x), x)\, dx\, dF(\theta) > 0, \tag{41}$$

and define

$$p_T^*(\theta) \equiv v(\tilde{q}(\theta), \theta) - \int_{\underline{\theta}}^{\theta} v_\theta(\tilde{q}(x), x)\, dx + k, \quad p_N^*(\theta) \equiv k. \tag{42}$$

We first show that the mechanism $(\tilde{q}, p^*)$ satisfies incentive compatibility (9), that is, for all θ, θ':

$$v(\tilde{q}(\theta), \theta) - p_T^*(\theta) \;\; \geq \;\; -p_N^*(\theta'), \tag{43}$$

$$v(\tilde{q}(\theta), \theta) - p_T^*(\theta) \;\; \geq \;\; v(\tilde{q}(\theta'), \theta) - p_T^*(\theta'). \tag{44}$$

Inequality (43) is immediate by the definition of p^*. To see (44), let $\theta' > \theta$. By definition of p_T^*, the difference between the left and the right hand side of (44) is

$$-\int_{\theta}^{\theta'} v_\theta(\tilde{q}(x), x)\, dx - v(\tilde{q}(\theta'), \theta) + v(\tilde{q}(\theta'), \theta'). \tag{45}$$

Since $\tilde{q}(\cdot)$ is increasing and $v_{q\theta} > 0$, and since $\theta' > \theta$, we have

$$-\int_{\theta}^{\theta'} v_\theta(\tilde{q}(x), x)\, dx \geq -\int_{\theta}^{\theta'} v_\theta(\tilde{q}(\theta'), x)\, dx = -v(\tilde{q}(\theta'), \theta') + v(\tilde{q}(\theta'), \theta). \tag{46}$$

Thus, expression (45) is larger than 0, and this proves (44) for $\theta' > \theta$. The argument for $\theta' < \theta$ is analogous.

Next, we verify the participation constraint (10). We note in passing that (42) implies that $h(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) = 1$ for all $\theta \in \Theta$. Hence, $U(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) = v(\tilde{q}(\theta), \theta) - p_T^*(\theta)$. Thus,

$$\int_{\underline{\theta}}^{\bar{\theta}} U(q(\theta), p(\theta) \,|\, \theta)\, dF(\theta) \;\; = \;\; \int_{\underline{\theta}}^{\bar{\theta}} [v(\tilde{q}(\theta), \theta) - p_T^*(\theta)]\, dF(\theta) \tag{47}$$

$$= \;\; \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} v_\theta(\tilde{q}(x), x) dx\, dF(\theta) - k = 0, \tag{48}$$

where the last equality follows from the definition of k.

To complete the proof of (a), notice that $(\tilde{q}, p^*)$ is optimal, because the seller fully extracts the first–best surplus

$$\int_{\underline{\theta}}^{\bar{\theta}} \Pi(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) \, dF = \int_{\underline{\theta}}^{\bar{\theta}} [p_T^*(\theta) - c(\tilde{q}(\theta))] \, dF \tag{49}$$

$$= \int_{\underline{\theta}}^{\bar{\theta}} [v(\tilde{q}(\theta), \theta) - c(\tilde{q}(\theta))] \, dF - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} v_\theta(\tilde{q}(x), x) dx \, dF + k \tag{50}$$

$$= \tilde{S} - 0. \tag{51}$$

Thus, the seller can clearly not do better than this.

(b) It follows from (a) that under any optimal contract (q^*, p^*), the seller must fully extract the first–best surplus $\tilde{S}$. This implies that $q^*(\theta) = \tilde{q}(\theta)$ for almost all $\theta \in \Theta$. Since $q^* = \tilde{q}$ is strictly increasing in θ almost everywhere, incentive compatibility implies by a standard argument that $U(q^*(\theta), p^*(\theta) \,|\, \theta)$ is strictly increasing in θ almost everywhere. Moreover, full surplus extraction also implies that $h(q^*(\theta), p^*(\theta) \,|\, \theta) = 1$ almost everywhere. Therefore, there is a $\Theta' \in \mathcal{T}$ with $\int_{\Theta'} dF = 1$ such that $U(q^*(\theta), p^*(\theta) \,|\, \theta)$ is strictly increasing on Θ' and

$$U(q^*(\theta), p^*(\theta) \,|\, \theta) = v(q^*(\theta), \theta) - p_T^*(\theta) \geq -p_N^*(\theta) \quad \text{for all } \theta \in \Theta'. \tag{52}$$

We now prove the claim by showing that $v(q^*(\theta), \theta) - p_T^*(\theta) > -p_N^*(\theta)$ for all $\theta \in \Theta' \setminus \{\underline{\theta}\}$. Indeed, suppose to the contrary that there is a $\theta \in \Theta' \setminus \{\underline{\theta}\}$ with $v(q^*(\theta), \theta) - p_T^*(\theta) = -p_N^*(\theta)$. Because Θ' has mass 1 and $\theta > \underline{\theta}$, we can find a $\theta' \in \Theta'$ with $\theta > \theta'$. Thus, we have

$$U(q^*(\theta'), p^*(\theta') \,|\, \theta') < U(q^*(\theta), p^*(\theta) \,|\, \theta) = v(q^*(\theta), \theta) - p_T^*(\theta) = -p_N(\theta). \tag{53}$$

But this is a contradiction because incentive compatibility requires that $U(q^*(\theta'), p^*(\theta') \,|\, \theta') \geq -p_N(\theta)$ for type θ'. Q.E.D.

Proof of Proposition 2: (a) For given $\theta \in \Theta$, define

$$p_T^*(\theta) \equiv v(\tilde{q}(\theta), \theta), \qquad p_N^*(\theta) \equiv 0. \tag{54}$$

We first verify that $\tilde{q}(\cdot)$ satisfies the seller's no–commitment constraint (12). We show that the seller optimally selects $\tilde{q}(\cdot)$ given p^*. It follows from the definition of transfers in (54) that when the seller chooses $\tilde{q}(\theta)$, then the buyer is just willing to trade, and the seller's profit is $\Pi(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) = p_T^*(\theta) - c(\tilde{q}(\theta))$. Now consider a quality $q < \tilde{q}(\theta)$. Then the buyer exits and the seller's profit is

$$\Pi(q, p^*(\theta) \,|\, \theta) = p_N^*(\theta) - c(q) \leq 0 \leq v(\tilde{q}(\theta), \theta) - c(\tilde{q}(\theta)) = \Pi(\tilde{q}(\theta), p^*(\theta) \,|\, \theta), \tag{55}$$

where the second inequality follows from the assumption that the surplus from the first–best quality is positive. Consider next a quality $q > \tilde{q}(\theta)$. Then the buyer accepts to trade and the seller's profit is

$$\Pi(q, p^*(\theta) \,|\, \theta) = p_T^*(\theta) - c(q) < p_T^*(\theta) - c(\tilde{q}(\theta)) = \Pi(\tilde{q}(\theta), p^*(\theta) \,|\, \theta), \tag{56}$$

where the inequality follows because costs are strictly monotone in q. Thus, (55) and (56) establish that $\tilde{q}(\cdot)$ satisfies the no–commitment constraint (12).

Next, we verify the participation constraint (13). We note in passing that (54) implies that $h(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) = 1$ for all $\theta \in \Theta$. Hence, $U(\tilde{q}(\theta), p^*(\theta) \,|\, \theta) = v(\tilde{q}(\theta), \theta) - p_T^*(\theta) = 0$. Thus, the participation constraint is, in fact, binding.

To complete the proof of (a), notice that $(\tilde{q}, p^*)$ is optimal, because the participation constraint is binding so that the seller fully extracts the first–best surplus $\tilde{S}$.

(b) It follows from (a) that under any optimal contract (q^*, p^*), the seller must fully extract the first–best surplus $\tilde{S}$. Therefore, it must hold that $h(q^*(\theta), p^*(\theta) \,|\, \theta) = 1$ for almost all $\theta \in \Theta$. That is, there is $\Theta' \in \mathcal{T}$ with $\int_{\Theta'} dF = 1$ and $v(q^*(\theta), \theta) - p_T^*(\theta) \geq -p_N^*(\theta)$ for all $\theta \in \Theta'$. Now suppose that there is a $\theta' \in \Theta'$ such that $v(q^*(\theta'), \theta') - p_T^*(\theta') > -p_N^*(\theta')$. Then, when faced with buyer type θ', the seller could increase his profit by slightly reducing $q^*(\theta')$, a contradiction to the no–commitment constraint (12). Thus, we must have that $v(q^*(\theta), \theta) - p_T^*(\theta) = -p_N^*(\theta)$ for all $\theta \in \Theta'$, and this completes the proof. Q.E.D.

Proof of Lemma 1: By definition of $M^+(\theta)$ in (17) one has $h(q(m), p(m) \,|\, \theta) = 1$ for all $\theta \in T^+(m)$. Therefore, by (1), $v(q(m), \theta) - p_T(m) \geq -p_N(m)$ for all $\theta \in T^+(m)$. By continuity of $v(q, \cdot)$ this implies

$$v(q(m), \theta_\ell(m)) - p_T(m) \geq -p_N(m). \tag{57}$$

Now suppose that $v(q(m), \theta_\ell(m)) - p_T(m) > -p_N(m)$. Since $v_\theta(q, \theta) > 0$, this implies

$$v(q(m), \theta) - p_T(m) > -p_N(m) \quad \text{for all } \theta \in T^+(m). \tag{58}$$

But this means that all buyer types who buy quality $q(m)$ after reporting m would also purchase a quality slightly below $q(m)$. Hence the seller could gain by reducing $q(m)$, a contradiction to (14). Q.E.D.

Proof of Lemma 2: We first prove that $p_N(m) = p_N(m')$ if $T^+(m) \neq \emptyset$ and $T^+(m') \neq \emptyset$. By Lemma 1 there is a $\theta_\ell(m)$ and a $\theta_\ell(m')$ such that

$$\begin{aligned} U(q(m), p(m) \,|\, \theta_\ell(m)) &= v(q(m), \theta_\ell(m)) - p_T(m) = -p_N(m), \\ U(q(m'), p(m') \,|\, \theta_\ell(m')) &= v(q(m'), \theta_\ell(m')) - p_T(m') = -p_N(m'). \end{aligned} \tag{59}$$

Since $m \in M^+(\theta)$ for all $\theta \in T^+(m)$, one has $U(q(m), p(m) \,|\, \theta) \geq U(q(m'), p(m') \,|\, \theta)$ for all $\theta \in T^+(m)$ and $U(q(m'), p(m') \,|\, \theta) \geq U(q(m), p(m) \,|\, \theta)$ for all $\theta \in T^+(m')$. By continuity of $U(q, p \,|\, \cdot)$ this implies

$$\begin{aligned} U(q(m), p(m) \,|\, \theta_\ell(m)) &\geq U(q(m'), p(m') \,|\, \theta_\ell(m)), \\ U(q(m'), p(m') \,|\, \theta_\ell(m')) &\geq U(q(m), p(m) \,|\, \theta_\ell(m')). \end{aligned} \tag{60}$$

Further, by (2)

$$U(q(m'), p(m') \,|\, \theta_\ell(m)) \geq -p_N(m'), \quad U(q(m), p(m) \,|\, \theta_\ell(m')) \geq -p_N(m). \tag{61}$$

By (59)–(61), we have $-p_N(m) \geq -p_N(m')$ and $-p_N(m') \geq -p_N(m)$. Therefore $p_N(m) = p_N(m')$.

It remains to show that $\theta_\ell(m) = \theta_\ell(m')$ if $T^+(m) \neq \emptyset$ and $T^+(m') \neq \emptyset$. Suppose the contrary, i.e. $\theta_\ell(m) \neq \theta_\ell(m')$. Without loss of generality, let $\theta_\ell(m) < \theta_\ell(m')$. Note that $q(m) > 0$ and $q(m') > 0$ because $m \in M^+(\theta)$ for all $\theta \in T^+(m)$ and $m' \in M^+(\theta)$ for all $\theta \in T^+(m')$. Since each type reports optimally, it must be the case that

$$v(q(m'), \theta) - p_T(m') \geq v(q(m), \theta) - p_T(m) \quad \text{for all } \theta \in T^+(m'), \tag{62}$$

so that, by continuity of $v(q, \cdot)$,

$$v(q(m'), \theta_\ell(m')) - p_T(m') \geq v(q(m), \theta_\ell(m')) - p_T(m). \tag{63}$$

Since $\theta_\ell(m) < \theta_\ell(m')$ and $v_\theta(q, \theta) > 0$,

$$v(q(m), \theta_\ell(m')) - p_T(m) > v(q(m), \theta_\ell(m)) - p_T(m). \tag{64}$$

By (63) and (64),

$$v(q(m'), \theta_\ell(m')) - p_T(m') > v(q(m), \theta_\ell(m)) - p_T(m). \tag{65}$$

Because $p_N(m) = p_N(m')$ this yields a contradiction to Lemma 1, which implies that

$$v(q(m'), \theta_\ell(m')) - p_T(m') = -p_N(m') = -p_N(m) = v(q(m), \theta_\ell(m)) - p_T(m). \tag{66}$$

Q.E.D.

Proof of Proposition 3: (i) We first show that $\theta > \hat{\theta}$ implies $R(\theta) \subseteq M^+(\theta)$, i.e. all types $\theta > \hat{\theta}$ only select positive trade messages. If $T^+(m) = \emptyset$ for all $m \in M$, then $\hat{\theta} = \bar{\theta}$, and the claim trivially holds. So suppose there is an $m \in M$ such that $T^+(m) \neq \emptyset$. Contrary to the claim, suppose there is a $\theta > \hat{\theta}$ and an $m' \in M(\theta) \setminus M^+(\theta)$. Since reporting m' is optimal for type θ, we have:

$$\begin{aligned} U(q(m'), p(m') \mid \theta) &= \max[-p_T(m'), -p_N(m')] \\ &\geq v(q(m), \theta) - p_T(m) \\ &> v(q(m), \hat{\theta}) - p_T(m) \end{aligned} \tag{67}$$

The first line follows since $q(m')h(q(m'), p(m')|\theta) = 0$ and $v(0, \theta) = 0$, the second line follows since m' is optimal for type θ, and the third line follows since $\theta > \hat{\theta}$ and $v_\theta(q, \theta) > 0$. By definition of $\hat{\theta}$, we have $m \in M(\hat{\theta})$, i.e. reporting m is optimal for type $\hat{\theta}$. By (22) this implies

$$\begin{aligned} U(q(m), p(m) \mid \hat{\theta}) &= v(q(m), \hat{\theta}) - p_T(m) \geq U(q(m'), p(m') \mid \hat{\theta}) \\ &= \max[v(q(m'), \hat{\theta}) - p_T(m'), -p_N(m')] \\ &\geq \max[-p_T(m'), -p_N(m')] \end{aligned} \tag{68}$$

Thus, $v(q(m), \hat{\theta}) - p_T(m) \geq \max[-p_T(m'), -p_N(m')]$, which yields a contradiction to (67). This proves that $\theta > \hat{\theta}$ implies $R(\theta) \subseteq M^+(\theta)$.

To complete the proof of (i), it remains to show that there is a $\hat{q}$ such that $q(m) = \hat{q}$ for all $m \in R(\theta)$ with $\theta > \hat{\theta}$. Suppose the contrary. Then there is a $\theta > \hat{\theta}$ and a $\theta' > \hat{\theta}$ such that $m \in R(\theta)$ and $m' \in R(\theta')$ with $q(m) < q(m')$. Since we have shown above that $R(\theta) \subseteq M^+(\theta)$ and $R(\theta') \subseteq M^+(\theta')$, we know that $T^+(m) \neq \emptyset$ and $T^+(m') \neq \emptyset$. Hence, by (22)

$$v(q(m), \hat{\theta}) - p_T(m) = v(q(m'), \hat{\theta}) - p_T(m'). \tag{69}$$

Therefore, $q(m') > q(m)$, $\theta > \hat{\theta}$, and $v_{q\theta}(q, \theta) > 0$ implies

$$v(q(m'), \theta) - v(q(m), \theta) > v(q(m'), \hat{\theta}) - v(q(m), \hat{\theta}) = p_T(m') - p_T(m). \tag{70}$$

But this is a contradiction because $v(q(m), \theta) - p_T(m) \geq v(q(m'), \theta) - p_T(m')$ as $m \in R(\theta)$.

(ii) Let $\theta < \hat{\theta}$ and suppose to the contrary that there is an $m \in R(\theta)$ such that $q(m) > 0$ and $h(q(m), p(m)|\theta) = 1$. This implies

$$v(q(m), \theta) - p_T(m) \geq -p_N(m). \tag{71}$$

Since $m \in R^+(\theta)$, it follows by definition that $T^+(m) \neq \emptyset$. Hence (22) holds for m and $p_N(m) = \hat{p}_N$. However, since $\theta < \hat{\theta}$ and $v_\theta(q, \theta) > 0$, (71) and (22) contradict each other. Q.E.D.

Proof of Lemma 3: For arbitrary (q', θ') with $v(q', \theta') - p_T = -p_N$, define

$$\phi(q', \theta'; \hat{\theta}) = \int_{\hat{\theta}}^{\bar{\theta}} \frac{\Pi(q', t \mid \theta)}{1 - F(\hat{\theta})} dF(\theta). \tag{72}$$

With the definition of Π in (3) we have

$$\phi(q', \theta'; \hat{\theta}) = \begin{cases} p_T - c(q') & \text{if} \quad v(q', \hat{\theta}) - p_T > -p_N, \\ \frac{F(\theta') - F(\hat{\theta})}{1 - F(\hat{\theta})} p_N + \frac{1 - F(\theta')}{1 - F(\hat{\theta})} p_T - c(q') & \text{if} \quad v(q', \hat{\theta}) - p_T \leq -p_N. \end{cases} \tag{73}$$

Hence, ϕ is strictly decreasing in q' if $v(q', \hat{\theta}) > p_T - p_N$, or equivalently if $\theta' < \hat{\theta}$. Therefore, any maximizer (q, θ) of $\phi(q', \theta'; \hat{\theta})$ satisfies $v(q, \theta) - p_T = -p_N$ and $\theta \geq \hat{\theta}$, which are the constraints in (29). Hence, since the bottom line on the right hand side in (73) coincides with the objective in (28):

$$(q, \theta) \in \underset{(q', \theta')}{\operatorname{argmax}} \, \phi(q', \theta'; \hat{\theta}) \quad \Leftrightarrow \quad (q, \theta) = (q^*(\hat{\theta}, p), \theta^*(\hat{\theta}, p)). \tag{74}$$

Consequently: $(\hat{q}, \hat{\theta})$ satisfies (24) and (25) $\Leftrightarrow$ $(\hat{q}, \hat{\theta}) \in \operatorname{argmax}_{(q', \theta')} \phi(q', \theta'; \hat{\theta})$ $\Leftrightarrow$ $(\hat{q}, \hat{\theta}) = (q^*(\hat{\theta}, p), \theta^*(\hat{\theta}, p))$ $\Leftrightarrow$ $(\hat{q}, \hat{\theta})$ satisfies (30). Q.E.D.

Proof of Lemma 4: For given $\hat{\theta}$, and for all (q', θ') denote the objective in (28) by

$$\psi(q', \theta'; \hat{\theta}) = \frac{F(\theta') - F(\hat{\theta})}{1 - F(\hat{\theta})} p_N + \frac{1 - F(\theta')}{1 - F(\hat{\theta})} p_T - c(q'). \tag{75}$$

Below, we show that under convexity of F, ψ is concave in (q', θ') for all θ. Moreover, since $p_T > p_N$, it is evident that ψ is decreasing in q' and θ'. Further, since v is quasi–concave by assumption, the constraint $v(q', \theta') = p_T - p_N$ describes a convex curve in (q, θ)-space. These three observations imply that the necessary first–order conditions for problem (28) are already sufficient to deliver a maximum. That is, $(\hat{q}, \hat{\theta}) = (q^*(\hat{\theta}, p), \theta^*(\hat{\theta}, p))$ if and only if the following Kuhn-Tucker conditions hold:

$$\psi_{q'}(\hat{q}, \hat{\theta}; \hat{\theta}) - \lambda v_q(\hat{q}, \hat{\theta}) = 0, \tag{76}$$

$$\psi_{\theta'}(\hat{q}, \hat{\theta}; \hat{\theta}) - \lambda v_\theta(\hat{q}, \hat{\theta}) \leq 0, \tag{77}$$

$$v(\hat{q}, \hat{\theta}) = p_T - p_N, \tag{78}$$

for some $\lambda \in \mathbb{R}$. It is easy to see that these conditions are equivalent to (31) and (32). This establishes the equivalence between (30) on the one and (31) and (32) on the other hand.

It remains to show that ψ is concave in (q', θ'). Observe first that the cross-partials $\psi_{q'\theta'}$ are zero. Thus, ψ is concave if and only if $\psi_{q'q'}$ and $\psi_{\theta'\theta'}$ are each negative. We have:

$$\psi_{q'q'}(q', \theta'; \theta) = -c''(q') \quad \text{and} \quad \psi_{\theta'\theta'}(q', \theta'; \theta) = -\frac{F''(\theta')}{1 - F(\theta)}(p_T - p_N). \tag{79}$$

Since $c'' > 0$ by assumption, the left expression is negative. Further, since $p_T - p_N > 0$, a sufficient condition for the right expression to be negative is that $F'' \geq 0$, and this completes the proof. Q.E.D.

Proof of Proposition 4: Insert $p_T - p_N$ from (32) in (31), and observe that the seller's problem becomes thus independent of transfers. This yields the claim. Q.E.D.

Proof of Lemma 5: Consider some continuation game $\Gamma(\bar{m}, \bar{q})$ and suppose that (40) does not hold. Then there exists an $\varepsilon \in (0, \bar{q})$ such that for all $\theta \in \hat{T}(\bar{m}, \bar{q})$

$$U(\bar{q} - \varepsilon, m^*(\theta), z_S^*(\theta), z_B^*(\theta)|\theta) \geq U(\bar{q} - \varepsilon, m^*(\theta), z_S^*(\theta), z'_B|\theta) \quad \text{for all } z'_B \in Z_B. \tag{80}$$

Let $\tilde{T}(\bar{m}) = \{\theta \in \Theta \,|\, \bar{m} = m^*(\theta)\}$. Note that $\hat{T}(\bar{m}, \bar{q}) \subseteq \tilde{T}(\bar{m})$. By (80), for each type $\theta \in \hat{T}(\bar{m}, \bar{q})$ the message $z_B^*(\theta)$ is a best response against $z_S^*(\theta)$ also in the continuation game $\Gamma(\bar{m}, \bar{q} - \varepsilon)$. Now consider $\theta \in \tilde{T}(\bar{m}) \setminus \hat{T}(\bar{m}, \bar{q})$. For these types $x^*(\theta) = 0$ in the continuation game $\Gamma(\bar{m}, \bar{q})$, i.e. they do not trade after observing $\bar{q}$. Thus it is also optimal for them not to buy $\bar{q} - \varepsilon$. This immediately implies that also for these types $z_B^*(\theta)$ remains a best response against $z_S^*(\theta)$ in the continuation game $\Gamma(\bar{m}, \bar{q} - \varepsilon)$.

Note that by (36), the seller's best response in the message game following the choice of q does not depend on the actual choice of q. Therefore, $z_S^*(\theta)$ remains a best response against $z_B^*(\cdot)$ also in the continuation game $\Gamma(\bar{m}, \bar{q} - \varepsilon)$. This together with argument in the previous paragraph proves that $(z_S^*, z_B^*(\cdot))$ is an equilibrium in the continuation game $\Gamma(\bar{m}, \bar{q} - \varepsilon)$.

Hence, Condition 1 implies that for all BNE $(\tilde{z}_S, \tilde{z}_B(\cdot)) \in E(\bar{m}, \bar{q} - \varepsilon)$ it holds that

$$p(\bar{m}, z_S^*, z_B^*(\theta)) = p(\bar{m}, \tilde{z}_S, \tilde{z}_B(\theta)) \tag{81}$$

for almost all $\theta \in \tilde{T}(\bar{m})$. Thus, the seller gets the same expected payment in $\Gamma(\bar{m}, \bar{q} - \varepsilon)$ as he gets in $\Gamma(\bar{m}, \bar{q})$. Since the quality $\bar{q} - \varepsilon$ is less costly than $\bar{q}$, the seller would therefore be better off by choosing quality $\bar{q} - \varepsilon$ at stage 2. This contradicts the condition that in equilibrium $\bar{q}$ maximizes the seller's profit after receiving the message $\bar{m}$.

It remains to show that $x(m^*(\theta), z_S^*(\theta), z'_B) = 0$ if (40) holds. Suppose the contrary. Then $x(m^*(\theta), z_S^*(\theta), z'_B) = x^*(\theta) = 1$ and (40) implies that

$$p(m^*(\theta), z_S^*(\theta), z'_B) < p(m^*(\theta), z_S^*(\theta), z_B^*(\theta)). \tag{82}$$

Therefore

$$U(\bar{q}, m^*(\theta), z_S^*(\theta), z_B'|\theta) = v(\bar{q}, \theta) - p(m^*(\theta), z_S^*(\theta), z_B') > \\ v(\bar{q}, \theta) - p(m^*(\theta), z_S^*(\theta), z_B^*(\theta)) = U(\bar{q}, m^*(\theta), z_S^*(\theta), z_B^*(\theta)|\theta). \tag{83}$$

This means that in the continuation game $\Gamma(\bar{m}, \bar{q})$ type θ would gain by deviating to $z_B' \neq z_B^*(\theta)$, a contradiction to equilibrium condition (37). Q.E.D.

Proof of Proposition 5 (i) Consider qualities $0 < q_1^* < q_2^*$ and define the sets

$$\Theta_i = \{\theta \in \Theta \mid q^*(\theta) = q_i^* \wedge x^*(\theta) = 1\}, \quad i = 1, 2. \tag{84}$$

We first proof the following:

$$\Theta_2 \neq \emptyset \quad \Rightarrow \quad \Theta_1 \text{ has measure zero.} \tag{85}$$

Suppose to the contrary that $\Theta_2 \neq \emptyset$ and Θ_1 has positive measure. Note that for all $\theta \in \Theta_i$, the equilibrium payments have to be the same: $p(m^*(\theta), z_S^*(\theta)), z_B^*(\theta)) = p_i^*$. (Otherwise, if there were two types in Θ_i with different equilibrium payments, then since the chosen quality is constant on Θ_i, the type with the higher payment could gain by sending the messages m and z_B of the type with the lower payment.) Therefore, since no type in Θ_i has an incentive to mimic the equilibrium strategy of a type in Θ_j, we have for all $\theta_i \in \Theta_i$:

$$v(q_i^*, \theta_i) - p_i^* \geq v(q_j^*, \theta_i) - p_j^*. \tag{86}$$

Next, since Θ_1 has positive measure, Θ_1 contains at least two types $\theta_1' < \theta_1''$. Let $m_1 = m^*(\theta_1')$ with $q(m_1) = q_1^*$ be the message used by the type θ_1'. Moreover, let m_2 be some message with $q(m_2) = q_2^*$ and $m_2 = m^*(\theta_2)$ for some $\theta_2 \in \Theta_2$. Further, define by $\theta_{\ell i} = \inf \hat{T}(m_i, q_i^*)$ the lowest type in Θ_i who uses the message m_i.

Since v is continuous in q, we can find for all $\rho > 0$ an $\varepsilon > 0$ such that $v(q_i^*, \theta_{\ell i}) - v(q_i^* - \varepsilon, \theta_{\ell i}) < \rho$. Further, by Lemma 5 we can find a $\tilde{\theta}_i \in \hat{T}(m_i, q_i)$ and a message $z_B^i(\rho, \varepsilon)$ such that $v(q_i^* - \varepsilon, \tilde{\theta}_i) - p_i^* < -p(m_i, z_S^*(\tilde{\theta}_i), z_B^i)$. Because v is increasing in θ, we have $v(q_i^* - \varepsilon, \theta_{\ell i}) \leq v(q_i^* - \varepsilon, \tilde{\theta}_i)$, and we therefore obtain

$$\begin{aligned} v(q_i^*, \theta_{\ell i}) - p_i^* &\leq v(q_i^*, \theta_{\ell i}) - v(q_i^* - \varepsilon, \theta_{\ell i}) + v(q_i^* - \varepsilon, \tilde{\theta}_i) - p_i^* \\ &< \rho - p(m_i, z_S^*(\tilde{\theta}_i), z_B^i). \end{aligned} \tag{87}$$

Moreover, in equilibrium no type $\theta_i \in \hat{T}(m_i, q_i)$ has an incentive to first report m_j and then z_B^j, $j \neq i$:

$$v(q_i^*, \theta_i) - p_i^* \geq -p(m_j, z_S^*(\tilde{\theta}_j), z_B^j) \quad \forall\, \theta_i \in \hat{T}(m_i, q_i). \tag{88}$$

By continuity of $v(q,\cdot)$, this inequality carries over to $\theta_{\ell i}$:

$$v(q_i^*,\theta_{\ell i}) - p_i^* \geq -p(m_j, z_S^*(\tilde{\theta}_j), z_B^j). \tag{89}$$

Combining (87) and (89) implies that for all $\rho > 0$:

$$\rho - p(m_1, z_S^*(\tilde{\theta}_1), z_B^1) > v(q_1^*,\theta_{\ell 1}) - p_1^* \geq -p(m_2, z_S^*(\tilde{\theta}_2), z_B^2), \tag{90}$$

$$\rho - p(m_2, z_S^*(\tilde{\theta}_2), z_B^2) > v(q_2^*,\theta_{\ell 2}) - p_2^* \geq -p(m_1, z_S^*(\tilde{\theta}_1), z_B^1). \tag{91}$$

Thus, if ρ goes to zero, a sandwich argument yields that

$$v(q_1^*,\theta_{\ell 1}) - p_1^* = v(q_2^*,\theta_{\ell 2}) - p_2^*. \tag{92}$$

Now, by continuity of $v(q,\cdot)$, the inequality (86) carries over to the types $\theta_{\ell i}$:

$$v(q_1^*,\theta_{\ell 1}) - p_1^* \geq v(q_2^*,\theta_{\ell 1}) - p_2^*, \qquad v(q_2^*,\theta_{\ell 2}) - p_2^* \geq v(q_1^*,\theta_{\ell 2}) - p_1^*. \tag{93}$$

Hence, (92) implies that $v(q_2^*,\theta_{\ell 2}) \geq v(q_2^*,\theta_{\ell 1})$ and $v(q_1^*,\theta_{\ell 1}) \geq v(q_1^*,\theta_{\ell 2})$. Consequently: $\theta_{\ell 1} = \theta_{\ell 2}$. Thus, since $q_1^* < q_2^*$, (92) together with $v_{q\theta} > 0$ implies that for the type $\theta_1'' > \theta_1' \geq \theta_{\ell 1}$, we have

$$v(q_2^*,\theta_1'') - v(q_1^*,\theta_1'') > p_2^* - p_1^*, \tag{94}$$

a contradiction to (86). This establishes (85).

We can now demonstrate (i). If $q^*(\cdot) = 0$ almost everywhere on Θ, the claim is true for $\hat{\theta} = \bar{\theta}$. So suppose there is a positive measure set $\Theta' \subseteq \Theta$ with $q^*(\cdot) > 0$ and $x^*(\cdot) = 1$ on Θ'. By incentive compatibility, $q^*(\cdot)$ and $x^*(\cdot)$ are (weakly) increasing on Θ. This implies first that $\sup \Theta' = \bar{\theta}$. Second, together with (85) it implies that $q^*(\cdot)$ is in fact constant on Θ' except possibly at $\hat{\theta} = \inf \Theta'$, and this implies (i).

(ii) Let (q^*, x^*) have the form described under (i). Let the exit option contract be given as follows: $M = \{m_\ell, m_h\}$, $Z_B = \{T, N\}$ and $Z_S = \emptyset$, and define $x(m,N) = 0$ and $x(m,T) = 1$ for all m. Moreover, let $p(m_\ell, N) = p(m_\ell, T) = 0$ and $p(m_h, N) = 0$ and $p(m_h, T) = v(\hat{q},\hat{\theta})$. It is easy to verify that (M, Z, x, p) implements (q^*, x^*). Q.E.D.

8 References

Aghion, P., M. Dewatripont, and P. Rey (1994): "Renegotiation Design with Unverifiable Information," *Econometrica*, 62, 257-282.

Bester, H., and R. Strausz (2001): "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica*, 69, 1077-1098.

Bester, H., and R. Strausz (2007): "Contracting with Imperfect Commitment and Noisy Communication," *Journal of Economic Theory*, 136, 236-259.

Che, Y.-K., and D. B. Hausch (1999): "Cooperative Investments and the Value of Contracting," *American Economic Review*, 89, 125-147.

Chung, T. (1991): "Incomplete Contracts, Specific Investments and Risk-Sharing," *Review of Economic Studies*, 58, 1031-1042.

Compte, O., and P. Jehiel (2007): "On Quitting Rights in Mechanism Design," *American Economic Review*, 97, 137-141.

Edlin, A., and B. Hermalin (2007): "Contract Renegotiation and Options in Agency Problems," *Journal of Law, Economics, and Organization*, 16, 395-423.

Edlin, A., and S. Reichelstein (1996): "Holdups, Standard Breach Remedies, and Optimal Investment," *American Economic Review*, 86, 478-501.

Evans, R. (2006): "Mechanism Design with Renegotiation and Costly Messages," University of Cambridge, UK, Cambridge Working Paper in Economics No. 0626.

Evans, R. (2008): "Simple Efficient Contracts in Complex Environments," *Econometrica*, 76, 459-491.

Grossman, S., and O. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Lateral and Vertical Integration," *Journal of Political Economy*, 94, 691-719.

Hart, O., and J. Moore (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755-785.

Hart, O., and J. Moore (1990): "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119-1158.

Hori, K. (2006): "Inefficiency in a Bilateral Trading Problem with Cooperative Investment," *The B.E. Journals in Theoretical Economics (Contributions)*, 6, 1-11.

Krishna, V., and J. Morgan (2008): "Contracting for Information under Imperfect Commitment," *Rand Journal of Economics*, 39, 905-925.

Maskin, E. (1977): "Nash Equilibrium and Welfare Optimality," MIT Working paper.

Maskin, E., and J. Moore (1999): "Implementation and Renegotiation," *Review of Economic Studies*, 66, 39-56.

Moore, J. (1992): "Implementation, Contracts, and Renegotiation in Environments with Complete Information" in *Advances in Economic Theory: Sixth World Congress*, vol. I, edited by J.–J. Laffont, Cambridge, Cambridge University Press.

Moore, J., and R. Repullo (1988): "Subgame Perfect Implementation," *Econometrica*, 56, 1191-1220.

Myerson, R. (1979): "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 47, 61-73.

Nöldeke, G., and K. Schmidt (1995): "Option Contracts and Renegotiation: A Solution to the Holdup Problem," *Rand Journal of Economics*, 26, 163-179.

Nöldeke, G., and K. Schmidt (1998): "Sequential Investments and Options to Own," *Rand Journal of Economics*, 29, 633-653.

Schmitz, P. W. (2002): "On the Interplay of Hidden Action and Hidden Information in Simple Bilateral Trading Problems," *Journal of Economic Theory*, 103, 444-460.

Zhao, R. (2008): "Rigidity in Bilateral Trade with Holdup," *Theoretical Economics*, 3, 85-121.