Exit Options in Incomplete Contracts with Asymmetric Information*

Helmut Bester[†]and Daniel Krähmer[‡]

March 28, 2012

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 deterministic 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, deterministic contracts are necessarily inefficient. The first–best can be achieved, however, through simple message games with stochastic terms of trade as off–equilibrium outcomes.

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

*We wish to thank Oliver Gürtler, Paul Heidhues, David Martimort, Tymofiy Mylovanov, Roland Strausz, Klaus Schmidt, an associate editor and two referees for helpful 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 Economics, 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 by the other party, 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.

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. The efficient level of quality is a strictly increasing function of the buyer's type. Quality is observable by the buyer, but we assume that it is publicly not verifiable (neither ex ante nor ex post). Consequently, quality cannot be legally enforced and so the seller has only imperfect commitment.

The basic insight of our paper is that efficient contracting must provide some kind of exit option for the buyer, which enables him to refuse trade after observing low quality. In environments with contractual incompleteness and symmetric information, the key role of exit clauses is well–established.[2] Our analysis thus extends this central role of exit clauses to environments with asymmetric information. This extension is non–trivial, because exit options that are efficient with symmetric information may fail to be incentive compatible when the buyer's type is his private information. Contracts with option clauses are often observed in practice. For example, contracts for house re–modeling, book publishing, advertising pilot campaigns, real estate agency services, or procurement contracts for specialized equipment frequently specify only payments contingent on delivery.[3] Also, performance contingent termination clauses in loan contracts or non–promotion clauses in labor contracts, or certain financial contracts such as convertible bond securities can be interpreted as forms of exit options.[4]

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 (1999), Moore

[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 (Grossman and Hart (1986) and Hart and Moore(1988)). We do not consider moral hazard or hidden actions, which can be observed only by the party who controls them.

[2]See, e.g., Chung (1991), Aghion, Dewatripont and Rey (1994), Nöldeke and Schmidt (1995, 1998), Edlin and Reichelstein (1996), Che and Chung (1999), Evans (2008).

[3]See Taylor(1993) or Che and Chung (1999).

[4]See Stiglitz and Weiss (1983), Kahn and Huberman (1988), Aghion, Bolton and Tirole (2004).

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.

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. It is well–known that, therefore, first–best efficiency can be attained in either of the two benchmark cases in which merely one of the imperfections is present. Indeed, when quality is contractible but the buyer's type is his private information, then since private information arrives *ex post*, the first–best can be achieved by a familiar screening contract. In the other benchmark case, when quality is non–contractible but the buyer's type is public information, 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.

When lack of verifiability and asymmetric information are combined, the simple exit option contract from the benchmark case will no longer achieve the first–best. To study implementability of the first–best, we consider the most general class of contracts, which allow the terms of trade – whether trade takes place or not and payments from the buyer and to the seller – to depend on a message by the buyer after he observes his type and on messages by both the seller and the buyer after quality has been chosen. Our main finding is that whether efficiency can be achieved or not depends crucially on the parties' ability to write contracts that for some messages specify trade of the good as a random event. If stochastic contracts are legally not enforceable so that the parties have to trade either with probability one or zero, we derive an impossibility result showing that efficiency losses are unavoidable. As we allow for payments to a third party, we obtain this result even without restricting contracts to be budget–balanced. However, if stochastic trading outcomes are contractually feasible, we show that the first–best can be achieved. In fact, as long as third party payments are not ruled out, a simple efficient exit option contract exists where, in contrast to the benchmark case, the buyer's exit option still induces trade with some positive probability. Moreover, we also provide an efficient mechanism when third party payments are impossible. This contract can be interpreted as a generalized exit option contract where the terms of exit for the buyer also depend on whether the seller, ex post, does or does not agree to trade.

To understand these results, it is useful to understand why efficient contracts can be designed in our two benchmark cases. When the buyer's type is private information and the seller can commit to quality, the efficient 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* than lower buyer types. On the other hand, when the buyer's valuation is public information, the efficient 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. Observe that the exit option contract would violate incentive compatibility when the buyer's type had to be elicited from the buyer. The reason is that under the exit option contract, any buyer type is indifferent between exit and trade. Moreover, the exit option has the same value for *any* buyer type because it only involves payments, but no trade. Thus, a lower buyer type could guarantee himself the same utility as a higher type by claiming to be the higher type and then exiting.

This tension in providing incentives jointly for the seller and the buyer also drives our inefficiency result for general deterministic contracts. To induce the seller to choose a given quality, the message game following the choice of quality must deter the seller from lowering quality. We demonstrate that this implies that the efficient contract would have to contain some form of exit option: there must be an exit–message inducing no trade, which the buyer chooses after a deviation from the given quality. This implies that if the seller has chosen the desired quality the buyer is essentially indifferent between his equilibrium message and the exit–message that he would play after a tiny deviation by the seller. Consequently, a low buyer type can guarantee himself approximately the same utility as a high buyer type by pretending to be a high type and then choosing the exit–message. However, since the first–best quality is increasing in the buyer's type, incentive compatibility would require that a low buyer type obtains a strictly smaller utility than a high buyer type.

This argument crucially rests on the fact that trade is deterministic. Because exit induces trade with probability zero, the buyer's utility from exit is independent of his type. This changes when stochastic trade is allowed. It then becomes possible to design a stochastic exit option which induces trade with a positive probability so that different buyer types value exit differently. This relaxes the incentive compatibility constraint and permits sorting of buyer types. In fact, we construct a stochastic exit option which induces the buyer to truthfully reveal his type and to trade if the seller selects the first–best quality after learning the buyer's type, and to exit if the seller shades quality. The remaining issue is then whether the seller can indeed be deterred from shading quality. This can be easily achieved if third party payments are possible, since then the seller can be fined harshly should the buyer exit. If third party payments are impossible, however, the payments to the seller after the buyer's choice of exit may not be small enough to deter the seller from shading quality. To resolve this problem, we construct a message game in which *both* parties become involved after the choice of quality. In this way we are able to implement budget–balanced contracts with sufficiently low payments after a quality deviation by the seller.

Interestingly, because the first–best requires trade to take place with probability one on the equilibrium path, in our context contractual randomization only serves to deter deviating behavior. The power of stochastic exit options results from creating incentives by specifying non–deterministic trade outcomes *off* the equilibrium path.[5]

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.[6] 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. As is well–known from the literature on implementation under complete information, message games may admit multiple equilibria. This implies that the seller's behavior on the equilibrium path may depend on a possibly implausible equilibrium selection in the message game off the path after a deviation by the seller. To resolve this problem, we adopt a version of the familiar concept of strong implementation, which requires all equilibria of the message game to induce the same trading outcome (see Maskin (1999)).

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 commit not to renegotiate ex post inefficient outcomes.[7] In contrast, the hold–

[5]This is quite different from Strausz (2006) who, in a context with adverse selection only, demonstrates that screening can be improved by making the actually implemented allocation stochastic.

[6]See the seminal papers by Maskin (1999) and Moore and Repullo (1988). For a survey, see Moore (1992).

[7]For implementation and renegotiation under complete information see Maskin and Moore (1999) or

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), Che and Sakovics (2004), Evans (2006, 2008)). Our paper is complementary to this debate. Rather than lack of commitment to enforce ex post inefficient default outcomes, we identify the parties' ability to write stochastic contracts as the central determinant of whether efficient contracting is possible or not.

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.[8] 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 the benchmark cases, where either quality is verifiable or the buyer's valuation is public information. Section 4 describes the contracting environment with private *and* unverifiable information. Section 5 derives an inefficiency result for optimal deterministic contracts, and Section 6 provides two first–best efficient mechanisms for stochastic contracts, one with and one without third party payments. Section 7 concludes. The proofs of all formal results are relegated to an appendix.

Segal and Whinston (2002).

[8]Hori (2006) and Zhao (2008) confirm Schmitz' result in mathematically more general environments. For their analysis of promotion schemes, Kahn and Huberman (1988) use a similar model where the employee privately invests in his productivity, which is only observed by the employer.

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 $t = 3$. The terms of trade specify a probability of trade x and payments p_B and p_S, where p_B denotes the payment made by the buyer and p_S denotes the payment received by the seller. We assume that $p_S \leq p_B$, that is, we rule out that the parties have access to external funds, but allow for the possibility that they transfer money to a third party. The parties' outside options are normalized to zero. 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 $c(q)$. The parties have quasi–linear utility so that given payments p_B and p_S the buyer's utility from trading q with probability x is $v(q, \theta)x - p_B$, and the seller's profit is $p_S - c(q)$. In stage $t = 3$ the buyer observes the seller's quality choice. Figure 1 summarizes the sequence of events.

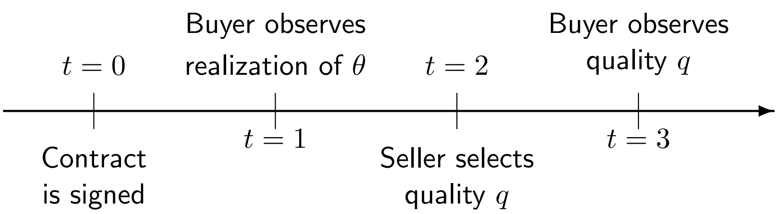

Figure 1: The Sequence of Events

The buyer's type θ is drawn from the finite set $\Theta = \{\theta_1, \ldots, \theta_i, \ldots, \theta_I\} \subset \mathbb{R}$, where without loss of generality $\theta_i < \theta_{i+1}$. Let $\gamma(\theta) > 0$ be the probability that type θ is realized. We make the following assumptions about $v(\cdot)$ and $c(\cdot)$:[9]

$$v(0, \theta) = 0,\ v_q(q, \theta) > 0, v_{qq}(q, \theta) \leq 0, \tag{1}$$

$$v(q, \theta) < v(q, \theta'),\ v_q(q, \theta) < v_q(q, \theta') \quad \text{for } \theta < \theta', \tag{2}$$

and

$$c(0) = 0,\ c'(q) > 0,\ c''(q) > 0. \tag{3}$$

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 and all $\theta \in \Theta$.

[9]Subscripts are used to denote partial derivatives.

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

$$\tilde{q}(\theta) \equiv \text{argmax}_{q \geq 0}\, v(q,\theta) - c(q) \tag{4}$$

is positive and unique. Also, by the last condition in (2), $\tilde{q}(\cdot)$ is strictly increasing in θ.

Our objective is to analyze what the parties can achieve by writing contracts. If, in addition to the trade probability x and the transfers p_B and p_S, the buyer and the seller were able to contractually specify the quality level $\tilde{q}(\theta)$ contingent upon the realization of θ, then always trading the first–best quality 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, 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 make the probability of trade and transfers contingent on communication between the parties. That is, the terms of trade can be contractually specified as functions of verifiable messages reported by the parties after each stage in which new information arrives. In Section 4, we describe in detail how the terms of trade can depend on the parties' communication. Before doing so, however, 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 analyze 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

$$\tilde{S} \equiv \sum\nolimits_{\theta} \left[v(\tilde{q}(\theta),\theta) - c(\tilde{q}(\theta))\right] \gamma(\theta). \tag{5}$$

Contractible q, asymmetric information about θ

Suppose quality q is verifiable so that the seller can contractually commit to a choice of quality. Consider a contract which requires the buyer to submit a report $\hat{\theta} \in \Theta$ after he has observed his type at stage 1. No further exchange of messages at later stages is required. The report $\hat{\theta}$ commits the seller to the quality $\tilde{q}(\hat{\theta})$, trade occurs with probability one, and the buyer pays $p(\hat{\theta}) = p_B(\hat{\theta}) = p_S(\hat{\theta})$ to the seller. It is easy to see that the payment

$$p(\hat{\theta}) \equiv c(\tilde{q}(\hat{\theta})) + p_0, \tag{6}$$

where p_0 is some constant, induces the buyer to reveal his type truthfully. Indeed, by (4),

$$v(\tilde{q}(\theta), \theta) - c(\tilde{q}(\theta)) \geq v(\tilde{q}(\theta'), \theta) - c(\tilde{q}(\theta')) \quad \text{for all} \quad \theta, \theta' \in \Theta. \tag{7}$$

Since the buyer learns his type only *ex post*, after the contract is signed, the seller can extract the full surplus by setting $p_0 \equiv \sum_\theta [v(\tilde{q}(\theta), \theta) - c(\tilde{q}(\theta))]\gamma(\theta)$. The next proposition summarizes this observation.

Proposition 1 *If q is contractible and θ private information, then there is a contract which implements the first–best.*

Note that for the incentive–compatibility payment schedule $p(\cdot)$, the buyer's *ex post* utility is strictly increasing in his type because

$$v(\tilde{q}(\theta), \theta) - p(\theta) \geq v(\tilde{q}(\theta'), \theta) - p(\theta') > v(\tilde{q}(\theta'), \theta') - p(\theta') \quad \text{for } \theta > \theta', \tag{8}$$

where the second inequality follows from (2).

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 other words, the choice of q is constrained by imperfect commitment on part of the seller. Consider a contract which gives the buyer an option at stage 3, after observing quality, to accept to trade or to exit, that is to refuse to trade. This means the buyer has two messages at stage 3, say T and E. The contract specifies that upon submitting message T trade occurs with probability one, while upon submitting message E there is no trade. The corresponding payments by the buyer to the seller are $p = p_S = p_B$ after reporting T, and $p^E = p^E_S = p^E_B$ after reporting E.

Now choose the payments so that the buyer is just indifferent between trade and no trade at the first–best quality:

$$v(\tilde{q}(\theta), \theta) - p = -p^E. \tag{9}$$

Then, at stage 3, the buyer optimally accepts any quality $q \geq \tilde{q}(\theta)$ and rejects any smaller quality. The seller's profit from choosing $q \geq \tilde{q}(\theta)$ is thus $p - c(q) = v(\tilde{q}(\theta), \theta) - c(q) + p^E$, and his profit from choosing $q < \tilde{q}(\theta)$ is p^E. Therefore, if p^E is set to zero, the seller effectively becomes the residual claimant. Consequently, he optimally chooses the first–best quality and extracts the first–best surplus. We summarize this observation in the next proposition.

Proposition 2 *If q is non–verifiable and θ public information, then there is a contract which implements the first–best.*

When the buyer's type is public information, the seller's commitment problem can be solved by offering the buyer an exit option at stage 3. An exit option implements the first–best quality only if the buyer is 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.[10]

Conditions (8) and (9) are incompatible to the extent that (8) requires all types to obtain a different utility *ex post*, while (9) implies that all types obtain the same utility, $-p^E$, *ex post*. This indicates a tension in providing appropriate incentives jointly for the buyer (incentive compatibility) and the seller (no–commitment). Thus, the question arises whether the first–best can still be implemented in the case when the seller cannot contractually commit to some quality and, at the same time, the buyer is privately informed about his type. We now turn to the contracting problem in this case.

4 Contracts

Since the parties observe new information during the course of the relation, the optimal contract requires the parties to report information after each stage in which they observe new information (see Myerson (1986)). We therefore consider the class of contracts which specify the terms of trade as functions of verifiable messages reported by the parties whenever they observe new information.[11] Thus, 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 at stage $t = 1$. In addition, a contract may require the parties to exchange

[10]Proposition 2 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.

[11]We restrict attention to face-to-face communication. As we argue below, this is without loss of generality since we focus on the question whether the first–best can be implemented.

messages after the seller's quality choice. That is, a contract may specify sets Z_S and Z_B of verifiable messages and require the seller and the buyer at stage 3, after having observed the chosen quality, to simultaneously select messages $z_S \in Z_S$ and $z_B \in Z_B$, respectively. Even though we describe the exchange of messages in stage $t = 3$ as a static game, this description may be thought of as the normal form representation of a dynamic game involving many stages of communication. Let $Z = Z_S \times Z_B$.

Thus, a contract (M, Z, x, p_S, p_B) consists of message sets M, Z and message contingent trade probabilities $x : M \times Z \to [0, 1]$ and message contingent payments $p_S : M \times Z \to \mathbb{R}$ and $p_B : M \times Z \to \mathbb{R}$, where p_S is a payment received by the seller, and p_B is a payment made by the buyer.[12] Note that we do not rule out payments from the seller or to the buyer, as p_S and p_B are not restricted to be non–negative.[13]

As indicated earlier, we allow for the possibility that the parties can make payments to a passive third party, but we assume that they do not have access to external funds, i.e. for all $(m, z) \in M \times Z$:

$$p_S(m, z) \leq p_B(m, z). \tag{10}$$

We say that a contract involves "third party payments" if this inequality is strict for some (m, z); if the equality holds for all (m, z), the contract is "budget–balanced". Allowing for third party payments is not uncontroversial since three party contracts of this sort may be difficult to implement, as they raise the problem of collusion between two of the agents against the third.[14] Even though third party payments may be impossible to enforce in some situations, we allow for them also for the methodological reason that this strengthens our impossibility result with deterministic trade below. Moreover, we will derive a possibility result with stochastic trade even when third party payments are impossible.

A contract induces an incomplete information game between the parties such that when in stage 1 the buyer reports m and in stage 3 the parties report $z = (z_S, z_B)$, trade occurs with probability $x(m, z)$, the seller receives $p_S(m, z)$, and the buyer has to pay $p_B(m, z)$. The buyer's and the seller's expected payoffs are defined as

$$U(m, q, z \mid \theta) = v(q, \theta)x(m, z) - p_B(m, z), \qquad \Pi(m, q, z) = p_S(m, z) - c(q). \tag{11}$$

[12] When the good is divisible, we may interpret $x \in (0, 1)$ as trade of only a fraction of the good.

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

[14] Cf. Hart and Moore (1988), especially footnote 20. For an argument in support of three party agreements, see Baliga and Sjöström (2009) who show in a complete information setting that if all coalitions have access to the same contracting technology, introducing a third party allows implementation of the first–best, even if the third party is corruptible.

It is easy to see that this environment entails the contracts of the benchmark cases discussed earlier as special cases.

Our aim is to study whether the first–best can be implemented in the class of contracts (M, Z, x, p_S, p_B). To formalize this, we now consider first contracts (M, Z, x, p_S, p_B) with arbitrary message sets M, Z_S, Z_B and state the restrictions imposed by the requirement that the parties play a Perfect Bayesian Equilibrium of the induced contracting game. Subsequently, we introduce a condition in the spirit of strong implementation that restricts the set of admissible contracts. We then state the conditions for first–best implementation.

Perfect Bayesian Equilibrium

Let the message sets M, Z_S, Z_B be arbitrary messages spaces and let $\mathcal{M}, \mathcal{Z}_{\mathcal{S}}, \mathcal{Z}_{\mathcal{B}}$ denote the respective σ-algebras on M, Z_S, Z_B. The functions x, p_S, p_B are taken to be measurable.

We denote the θ–type buyer's (pure) reporting strategy in stage 1 by $r(\theta) \in M$.[15] After receiving message m, the seller updates his beliefs about the buyer's type. We denote these beliefs by $\mu(T, m)$. Thus, upon observing message m, the seller believes that the buyer's true type is in the set $T \subseteq \Theta$ with probability $\mu(T, m)$. Given his beliefs, the seller chooses $q(m)$ to maximize his expected payoff.

After m and q have been selected in the previous stages, in stage 3 a continuation game starts in which the seller has imperfect information about the buyer's type. We refer to this game as $\Gamma(m, q)$. The seller enters $\Gamma(m, q)$ with the belief $\mu(\cdot, m)$. The game $\Gamma(m, q)$ is thus a (static) Bayesian game where a strategy for the seller is given by a function $\zeta_S(m, q)$ which specifies for each path (m, q) a message in Z_S, and a strategy for the buyer is a function $\zeta_B(\theta, m, q)$ which specifies a message in Z_B for each type θ and each path (m, q).

To constitute a Perfect Bayesian Equilibrium, the functions $(r, \mu, q, \zeta_S, \zeta_B)$ have to satisfy the following conditions: First, for each type θ and each path (m, q), the strategies (ζ_S, ζ_B) have to form a Bayesian Nash Equilibrium of $\Gamma(m, q)$. Thus, the seller's message $\zeta_S = \zeta_S(m, q)$ satisfies

$$\sum\nolimits_\theta p_S(m, \zeta_S, \zeta_B(\theta))\, \mu(\theta, m) \geq \sum\nolimits_\theta p_S(m, z_S, \zeta_B(\theta))\, \mu(\theta, m) \quad \text{for all } z_S \in Z_S, \tag{12}$$

and each buyer type θ with $m = r(\theta)$ selects a message $\zeta_B(\theta) = \zeta_B(\theta, m, q)$ such that

$$U(m, q, \zeta_S, \zeta_B(\theta) \mid \theta) \geq U(m, q, \zeta_S, z_B \mid \theta) \quad \text{for all } z_B \in Z_B. \tag{13}$$

[15] Here and in what follows, we restrict attention to pure strategies. As we argue below, this is without loss of generality for the purpose of our analyiss. See footnotes 16 and 17.

Note that in $t = 3$, the seller's production costs are already sunk so that in (12), he only cares about expected payments when choosing his message.

Second, the seller's choice of quality has to be optimal given his beliefs formed after receiving message m and given anticipated equilibrium play in stage 3. This means that $q(\cdot)$ has to satisfy the *no–commitment constraint*

$$q(m) \in \operatorname{argmax}_q \sum\nolimits_\theta \Pi(m, q, \zeta_S(m, q), \zeta_B(\theta, m, q))\ \mu(\theta, m) \quad \text{for all } m \in M. \tag{14}$$

Third, as the buyer anticipates that message m will induce the seller to select $q(m)$, and given anticipated equilibrium play in stage 3, he will select an optimal reporting strategy in stage 1. This means that $r(\cdot)$ has to satisfy the buyer's *communication incentive constraint*

$$r(\theta) \in \operatorname{argmax}_m U\left(m, q(m), \zeta_S(m, q(m)), \zeta_B(\theta, m, q(m)) \mid \theta\right) \quad \text{for all } \theta \in \Theta. \tag{15}$$

Finally, the seller's belief 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 $m = r(\theta)$ for some $\theta \in \Theta$. Of course, the belief μ determines the seller's choice of q also for messages that lie outside of the support of the buyer's reporting strategy. Yet, there are no consistency restriction on beliefs for such messages.

Admissible contracts

In stage 3, after the buyer has reported m and the seller has chosen $q(m)$, the parties are engaged in a message game. As is well–known, message games of this kind typically admit a multiplicity of equilibria. While some of these equilibria may implement the desired outcome, others may induce unintended outcomes. In particular, the literature on implementation in complete information environments deems mechanisms of the "shoot-the-liar" type implausible 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 undesirable Nash equilibria in which agents do not tell the truth.

In our case, the seller's equilibrium quality choice $q(m)$ is supported by the equilibrium outcome of the stage 3 message game $\Gamma(m, q')$ that would arise if the seller deviated from the equilibrium quality to some smaller quality q'. If the message game $\Gamma(m, q')$ has multiple equilibria, such as in "shoot-the-liar", the seller may be deterred from deviating from the equilibrium quality to q' through the selection of a (perhaps implausible, "bad") equilibrium in $\Gamma(m, q')$. Yet, if a different (perhaps more plausible) equilibrium was selected in $\Gamma(m, q')$, it could be that the seller's choice of $q(m)$ might no longer be optimal. To resolve this problem, we will apply a version of the familiar concept of *strong implementation*. In particular, for any message game $\Gamma(m, q')$ that starts after the seller has shirked, we require *all* equilibria of this game to have identical outcomes. This guarantees that the seller's quality choice is not driven by a possibly implausible equilibrium selection after a deviation by the seller from the equilibrium path.

More formally, we restrict the set of admissible contracts by imposing the following condition on a contract (M, Z, p, x).

Condition 1 If $(r, \mu, q, \zeta_B, \zeta_S)$ and $(r, \mu, q, \hat{\zeta}_B, \hat{\zeta}_S)$ are Perfect Bayesian Equilibria, then for all $q' < q(m)$:

$$x(m, \zeta_S(m, q'), \zeta_B(\theta, m, q')) = x(m, \hat{\zeta}_S(m, q'), \hat{\zeta}_B(\theta, m, q')), \tag{16}$$

$$p_S(m, \zeta_S(m, q'), \zeta_B(\theta, m, q')) = p_S(m, \hat{\zeta}_S(m, q'), \hat{\zeta}_B(\theta, m, q')), \tag{17}$$

$$p_B(m, \zeta_S(m, q'), \zeta_B(\theta, m, q')) = p_B(m, \hat{\zeta}_S(m, q'), \hat{\zeta}_B(\theta, m, q')) \tag{18}$$

for all θ with $m = r(\theta)$.

Thus, if the buyer type θ has reported $m \in M$ in stage $t = 1$ and the seller has deviated from the equilibrium quality $q(m)$ to a smaller quality q' in stage $t = 2$, Condition 1 implies that the trade outcome x and the payments p_S and p_B are uniquely determined by the outcome of the subsequent message game in $t = 3$, even when this game has multiple equilibria.[16]

First–best implementation

We now describe the conditions for first–best implementation. We say that a contract implements the first–best if there is a Perfect Bayesian Equilibrium of the induced contracting game satisfying Condition 1 such that, on the equilibrium path, the seller selects the first–best quality for all buyer types, trade takes place with probability one, and there are no third party payments.

[16] Condition 1 is stated for pure strategy equilibria (ζ_S, ζ_B) of the message game at stage 3. This is without loss of generality when we adopt Maskin's (1999, p.25) extension to implementation in mixed strategies which requires for mixed strategy equilibria of the message game that all outcomes be the same for all realizations in the support of the mixed strategies. This implies the special property that if there is a mixed strategy equilibrium in stage 3, then a player is indifferent between the messages in the support of his mixed strategy, *no matter* how the other player randomizes between the strategies in the support of his mixed strategy. In other words, any pure strategy profile in the support of the mixed strategy equilibrium forms a pure strategy equilibrium of the message game. Thus, to check Condition 1, it is sufficient to check it for all pure strategy equilibria and to show that there are no mixed strategy equilibria other than those given by mixture over the pure strategy equilibria.

Observe that since the first–best quality is different for every buyer type, it is necessary for first–best implementation that the buyer's message m in stage 1 truthfully reveals the buyer's type θ to the seller. It therefore follows by an argument in the spirit of the Revelation Principle that if the first–best can be implemented by some contract, it can be implemented by a "direct" contract where the messages space M coincides with the type space Θ, and that, moreover, induces the buyer to truthfully reveal his type in stage 1:[17]

$$r(\theta) = \theta \quad \text{for all } \theta \in \Theta. \tag{19}$$

This means that the seller's equilibrium belief is $\mu(\theta, \theta) = 1$ for all $\theta \in \Theta$.

In addition, first–best implementation requires that the seller selects the first–best quality after receiving message $m = r(\theta) = \theta$,

$$q(r(\theta)) = \tilde{q}(\theta) \quad \text{for all } \theta \in \Theta, \tag{20}$$

and that, on the equilibrium path, trade takes place with probability one and there are no third party payments, that is, for all $\theta \in \Theta$:

$$x(\theta, \zeta_S(\theta, \tilde{q}(\theta)), \zeta_S(\theta, \theta, \tilde{q}(\theta)) = 1, \tag{21}$$

$$p_S(\theta, \zeta_S(\theta, \tilde{q}(\theta)), \zeta_S(\theta, \theta, \tilde{q}(\theta)) = p_B(\theta, \zeta_S(\theta, \tilde{q}(\theta)), \zeta_S(\theta, \theta, \tilde{q}(\theta)). \tag{22}$$

Summing up, we say that the first–best is implementable if there is a contract (M, Z, x, p_S, p_B) with $M = \Theta$ so that there is a Perfect Bayesian Equilibrium $(r, \mu, q, \zeta_S, \zeta_B)$ of the induced contracting game satisfying Condition 1 and conditions (19) to (22).

We now turn to the question whether first–best implementation is possible or not. We first consider deterministic contracts in which the probability of trade is either zero or one. Studying deterministic contracts is interesting, because many real–world contracts are deterministic. For example, the benchmark contracts considered in Section 3 are deterministic. Moreover, in practice, random mechanisms are often infeasible. Indeed, Che and Hausch (1999) argue that random mechanism are questionable to implement because of legal enforcement problems. Our main result for deterministic contracts is an inefficiency result which shows that when trade is deterministic, the first–best cannot be implemented.

While deterministic contracts are an important benchmark, they are less appealing from a theoretical point of view. Also in many situations, random mechanisms can be enforced. For example, if the parties have access to legally binding arbitration, a trusted

[17] In particular, our restriction to pure reporting strategies for the buyer in stage 1 is without loss of generality.

mediator can perform the required randomization, and the court simply has to enforce the mediator's decision. Alternatively, for divisible goods stochastic contracts correspond to trade of a fraction of the good.[18] Therefore, we treat in a second step the case when the parties can commit to random trade. We show that allowing for random trade substantially enlarges the set of implementable outcomes. Indeed, we provide two explicit mechanisms that implement the first–best, one making use of third party payments, and the other respecting budget–balancedness not only on but also off the equilibrium path.

5 Deterministic contracts

In this section, we study deterministic contracts where the probability of trade is restricted to be zero or one: $x : M \times Z \to \{0, 1\}$. We demonstrate that it is impossible to implement the first–best. We derive necessary conditions for first–best implementation from the no–commitment and from the communication constraints and show that they are inconsistent with one another. If the first–best is implementable, then the buyer truthfully reveals his type θ in stage 1 and will then be offered the first–quality $\tilde{q}(\theta)$. Finally, in stage 3, the outcome of the message game determines whether trade takes place or not. To simplify notation, we write

$$\zeta_S^*(\theta) \equiv \zeta_S(\theta, \tilde{q}(\theta)), \quad \zeta_B^*(\theta) \equiv \zeta_B(\theta, \theta, \tilde{q}(\theta)), \tag{23}$$

$$x^*(\theta) \equiv x(\theta, \zeta_S^*(\theta), \zeta_B^*(\theta)), \quad p_S^*(\theta) = p_B^*(\theta) \equiv p_S(\theta, \zeta_S^*(\theta), \zeta_B^*(\theta)) \tag{24}$$

for the stage 3 equilibrium strategies, the trade probabilities, and the payments, respectively that are induced on the equilibrium path when the buyer is of type θ.[19]

The next lemma is key for the main result of this section. It says that any contract that implements the first–best necessarily features some exit option for the buyer. The existence of such an exit option is a consequence of the seller's no commitment constraint (14).

Lemma 1 *Suppose the first–best is implementable and that trade is deterministic. Then for all $\theta \in \Theta$ and all $\epsilon \in (0, \tilde{q}(\theta)]$, there is a message $z_B^E \in Z_B$ with $x(\theta, \zeta_S^*(\theta), z_B^E) = 0$ so that*

$$\begin{aligned} v(\tilde{q}(\theta) - \epsilon, \theta) - p_B^*(\theta) &< -p_B(\theta, \zeta_S^*(\theta), z_B^E) & (25)\\ &\leq v(\tilde{q}(\theta), \theta) - p_B^*(\theta). & (26)\end{aligned}$$

[18] We thank two anonymous referees for raising this issue.

[19] Note that on the equilibrium path, $p_S = p_B$ by (22).

The first inequality in Lemma 1 says that for any deviation of the seller by ϵ from the first–best quality, there is an exit-message z_B^E leading to no trade so that the buyer type θ prefers this message over accepting to trade the quality $\tilde{q}(\theta) - \epsilon$. The driving force behind this result is similar as in the benchmark case with non–contractible quality. Recall that in the benchmark case for an exit option contract to prevent the seller from shading quality, the buyer has to be left indifferent between trade and exit. Lemma 1 extends this insight to the general contracting environment. The message z_B^E creates a credible threat for the buyer that deters the seller from lowering quality ex post. It serves a similar function as the exit option in the benchmark case in restraining the seller's limited commitment.

The second inequality in Lemma 1 says that the type θ is (weakly) better off when he trades the quality $\tilde{q}(\theta)$ rather than choosing the exit–message z_B^E in stage 3. This follows directly from the equilibrium condition that under first–best implementation trading $\tilde{q}(\theta)$ is, by definition, optimal for the type θ.

Jointly, the two inequalities in Lemma 1 imply that the type θ is approximately indifferent between trade and choosing the exit–message z_B^E in stage 3. Moreover, observe that the right hand side of (25) depends only on the buyer's message θ, but is independent of his true type. Therefore, by submitting the message θ in stage 1 and then an appropriate exit–message z_B^E in stage 3, *any* type $\theta' \neq \theta$ can secure himself almost the same utility as the equilibrium utility of the type θ.

However, this is inconsistent with the buyer's communication incentive constraint which, as in the benchmark case with only asymmetric information, implies that a buyer type $\theta > \theta'$ must get a strictly larger utility than the buyer type θ' under first–best implementation. Therefore, we obtain the following impossibility result for deterministic trade.

Proposition 3 *Suppose that trade is deterministic. Then the first–best cannot be implemented.*

Proposition 3 confirms the intuition from the benchmark cases that it is impossible to provide incentives jointly for the buyer to reveal his type and for the seller to choose the first–best quality. By Propositions 1 and 2, this inefficiency result is driven by the combined presence of private information and contractual completeness.

It is noteworthy that our impossibility result obtains even though the parties have rather powerful contracting possibilities to the extent that they can commit ex ante to the terms of trade and, in particular, can implement inefficient ex post outcomes by leaving the good or money on the table. Moreover, the impossibility result does not depend on the fact that the parties communicate face-to-face in stage 1. It still holds true if the parties could communicate through a mediator a la Myerson (1986) in the first stage. This is so since first–best implementation requires the type of the buyer to be perfectly revealed. Consequently, any attempt by the mediator to only partially reveal the buyer's type so as to mitigate the seller's commitment problem would imply some inefficiency. On a related note, the argument behind Proposition 3 only exploits the fact that the first–best quality schedule is increasing in the buyer's type and therefore implementing it requires that the buyer's type is revealed. The same logic implies that it is impossible to implement *any* quality schedule that is increasing in the buyer's type. Hence, pooling of buyer types is unavoidable.[20]

6 Stochastic contracts

We now turn to the case when the parties can commit to stochastic trade: $x\colon M \times Z \to [0, 1]$. As the main result of this section, we explicitly construct two mechanisms that implement the first–best. Note that under the first–best, trade takes place with probability one on the equilibrium path. Therefore, what improves the power of stochastic over deterministic contracting is that it allows the parties to create incentives by specifying non–deterministic trade probabilities off the equilibrium path. Recall that with deterministic trade, first–best implementation fails because by Lemma 1 an exit–message leading to no trade at stage 3 is needed to deter the seller from shirking. But given the existence of such a message, it is impossible to sort types, because a low type can secure himself almost the same utility as the equilibrium utility of a high type by imitating this type at stage 1 and then choosing the exit–message at stage 3. This argument crucially relies on the fact that trade is deterministic: because exit leads to trade with probability 0, the buyer's utility from exit is independent of the buyer's type. This changes when stochastic trade is allowed. It then becomes possible to design an exit–message which induces trade with a positive probability so that different buyer types value exit differently. This relaxes the communication incentive constraint and permits a finer sorting of types. The question then remains if exit–messages can be found which achieve the dual goal of sorting types and deterring the seller from shading quality.

To construct a mechanism that implements the first–best, we proceed in two steps. We first consider a simple contract with a stochastic exit option for the buyer and demonstrate

[20]In particular, this applies to the optimal, second–best contract. Note that finding the optimal, second–best contract is not straightforward, because for this type of problem, it is well–known that it may be quite restrictive to use a direct face–to–face communication mechanism that induces truthful revelation. For example, as shown in Bester and Strausz (2001), an indirect mechanism may support outcomes that cannot be replicated by a direct mechanism. Further, communication may be improved by using a mediatior (see Myerson (1986) and Bester and Strausz (2007)).

that there is a contract in that class which satisfies the buyer's incentive constraints for first–best implementation. In the second step, we extend the simple contract to also provide incentives for the seller not to shirk. We show that the seller's incentives can easily be satisfied by a payment scheme that involves third party payments off the path. More interestingly, we also construct a contract which provides seller incentives and balances the budget even off the path.

In the first step, we focus only on the buyer's incentives and ignore the seller's incentives. We consider a contract with a stochastic exit option for the buyer. The contract requires the buyer to announce a type $m = \theta \in \Theta$ in stage 1, and then allows him to announce in stage 3 whether to trade the observed quality with probability 1 at the price $p_B(\theta)$, or to exit, which means that trade occurs with probability $x^E(\theta) \in [0, 1)$ and the buyer pays $p_B^E(\theta)$. In the first step, there are no seller messages at stage 3. We now state the incentive constraints so that the buyer announces his type truthfully in stage 1, trades the first–best quality in stage 3, and chooses exit for all qualities that fall below the first–best quality. First, in stage 3, the buyer trades if the chosen quality is the first–best, and exits otherwise:

$$\begin{aligned} v(\tilde{q}(\theta), \theta) - p_B(\theta) &\geq v(\tilde{q}(\theta), \theta)x^E(\theta) - p_B^E(\theta) \qquad \text{for all } \theta, & (27)\\ v(q, \theta)x^E(\theta) - p_B^E(\theta) &> v(q, \theta) - p_B(\theta) \qquad \text{for all } q < \tilde{q}(\theta). & (28)\end{aligned}$$

Second, in stage 1, the buyer has an incentive to announce his type truthfully, i.e. he must not gain by misrepresenting his type and then trade or exit in stage 3:

$$\begin{aligned} v(\tilde{q}(\theta), \theta) - p_B(\theta) \geq \max[\, & v(\tilde{q}(\theta'), \theta) - p_B(\theta'), \\ & v(\tilde{q}(\theta'), \theta)x^E(\theta') - p_B^E(\theta')\,] \text{ for all } \theta' \neq \theta. \qquad (29)\end{aligned}$$

The next lemma shows that the incentive constraints for the buyer can be satisfied.

Lemma 2 *For some constant k, define*

$$p_B(\theta) \equiv c(\tilde{q}(\theta)) + k. \tag{30}$$

Then there exist $x^E(\theta) \in (0, 1)$ and $p_B^E(\theta)$ with

$$p_B^E(\theta) < p_B(\theta) \tag{31}$$

so that (27), (28), and (29) hold.

The construction of the exit option in Lemma 2 is straightforward. By letting q go to $\tilde{q}(\theta)$, the conditions (27) and (28) pin down $p_B^E(\theta)$ as a function of $x^E(\theta)$ and $p_B(\theta)$. We then have to choose $x^E(\theta)$ and $p_B(\theta)$ so that (29) holds.

We now turn to the seller's limited commitment constraint. If the payment to the seller in case the buyer exits, p_S^E, is sufficiently small, the threat of exit deters the seller from lowering quality below the first–best. More precisely, recall that first–best implementation requires no third party payments on the path so that $p_S(\theta) = p_B(\theta)$. Now, let

$$p_S^E(\theta) < p_S(\theta) - c(\tilde{q}(\theta)). \tag{32}$$

Then when facing buyer type θ, the seller is better off by choosing the first–best quality than by choosing any smaller quality $q < \tilde{q}(\theta)$ and triggering exit by the buyer.[21] Thus, we have the following possibility result.

Proposition 4 *Let p_B, p_B^E and x^E satisfy (30) and (31) in Lemma 2. Further, let $p_S = p_B$, and let p_S^E satisfy (32) and $p_S^E \leq p_B^E$. Then the following contract implements the first–best:*

– *After the buyer announces a type and the seller chooses a quality, the buyer has two messages, "trade" and "exit", and the seller has no message in stage 3.*

– *The buyer's trade–message induces trade with probability 1, and the buyer pays $p_B(\theta) = p_S(\theta)$ to the seller.*

– *The buyer's exit–message induces trade with probability $x^E(\theta)$, the seller receives $p_S^E(\theta)$ and the buyer pays $p_B^E(\theta)$.*

Proposition 4 shows that, once we allow for stochastic trade and third party payments, the first–best can be implemented by a "simple" exit option contract in the spirit of the exit option contracts in the benchmark cases. While the importance of exit options in overcoming contracting limitation caused by non–verifiability is well–appreciated, the result extends the central role of exit options to environments where there is also asymmetric information, providing a further rationale for the prevalence of exit options in practice.

A limitation of Proposition 4 is that it may require the parties to make payments to a third party once the buyer chooses exit in stage 3. As argued earlier, this is problematic when the possibility of collusion between the two agents against the third cannot be ruled out. We therefore now investigate the question of whether the first–best can be implemented even if third party payments are not possible.

To illustrate the problem when third party payments are impossible, consider again the payments in Lemma 2. Balancing the budget also after the buyer exits would require

[21] We thank an anonymous referee for indicating that with third party payments, the seller's incentives can be easily satisfied.

x, p	$z_B = N$	$z_B = E$	$z_B = T$
$z_S = T$	$0, k - \epsilon$	$x^E(\theta),\ \ p^E(\theta)$	$1, p(\theta)$
$z_S = N$	$0, k$	$0,\ \ \max[p^E(\theta),\ k] + \epsilon$	$0, \bar{p}(\theta)$

Figure 2: Contract terms (x, p) as a function of messages

that $p_S^E = p_B^E$. However, similarly to the logic with deterministic trade, p_B^E must be sufficiently large to prevent low buyer types from mimicking high buyer types in stage 1 and then exiting in stage 3. Therefore, when the parties cannot transfer money to a third party, the payment $p_S^E = p_B^E$ to the seller is in general not sufficiently small to deter the seller from shading quality.

To resolve this problem we to introduce further stage 3 messages, also for the seller, and use the buyer's exit–message to trigger an unraveling process that leads to no trade with probability one as the final outcome after a quality deviation by the seller. Intuitively, we give each player an additional no–trade–message and construct the terms of trade so that if the seller shaded quality and wanted to trade, the buyer would choose exit as a best response, yet given exit by the buyer, the seller's best response would be no–trade, and finally, given no–trade by the seller, the buyer's best response would be no–trade. This sequence of best responses eventually leads to a *unique* equilibrium after the seller has deviated from the first–best quality. In this equilibrium, the payments are low enough to deter the seller from shirking.

More precisely, suppose that after the buyer has announced θ in stage 1, the message game in stage 3 has three messages for the buyer, $Z_B = \{N, E, T\}$, and two messages for the seller, $Z_S = \{T, N\}$. Since we now rule out third party payments, the contract needs to specify only a single payment $p(m, z_S, z_B)$ from the buyer to the seller. Specifically, consider the contract with the trading-payment outcomes (x, p) as displayed in Figure 2, where $p(\theta), p^E(\theta)$ and $x^E(\theta)$ satisfy Lemma 2 where the subindex B is omitted. Moreover, $\epsilon > 0$ is sufficiently small and so by (30) and (31) we can choose the payment $\bar{p}(\theta)$ to satisfy

$$\max[p^E(\theta), k] + \epsilon < \bar{p}(\theta) < p(\theta). \tag{33}$$

When the buyer reports truthfully in stage 1, and the seller selects quality q in stage 2, the contract induces the game depicted in Figure 3 at stage 3.

We now argue that when the seller chooses a quality lower than first–best, $q < \tilde{q}(\theta)$, the unique equilibrium of the message game is (N, N). Thus, the contract satisfies Condition 1. To see that (N, N) is the unique equilibrium, observe that (N, N) is the only strategy pair that survives the iterated elimination of strictly dominated strategies. This, in

	$z_B = N$	$z_B = E$	$z_B = T$
$z_S = T$	$k-\epsilon,\ -k+\epsilon$	$p^E(\theta),\ v(q,\theta)x^E(\theta) - p^E(\theta)$	$p(\theta),\ v(q,\theta) - p(\theta)$
$z_S = N$	$k,\ -k$	$\max[p^E(\theta),\ k] + \epsilon,\ -\max[p^E(\theta),\ k] - \epsilon$	$\bar{p}(\theta),\ -\bar{p}(\theta)$

Figure 3: Message game when the seller selects q

particular, implies that there are no mixed strategy equilibria. Indeed, note that $z_B = T$ is strictly dominated by $z_B = E$ for the buyer by (28) and the first inequality in (33). Moreover, once $z_B = T$ is eliminated, $z_S = T$ is strictly dominated by $z_S = N$ for the seller. Since the buyer strictly prefers the outcome under (N, N) over (N, E), only the strategy pair (N, N) survives.

If, however, the seller does select the first–best quality $q = \tilde{q}(\theta)$, then by (27), T becomes a best response to $z_S = T$ for the buyer. Furthermore, by the second inequality in (33), the seller's best response to $z_B = T$ is T so that the messages (T, T) constitute an equilibrium. Therefore, the seller's profit from choosing $q = \tilde{q}(\theta)$ in stage 2 is $p(\theta) - c(\tilde{q}(\theta)) = k$. This means that he cannot gain from shirking.[22] We now state our efficiency result for budget–balanced contracts:

Proposition 5 *Consider the following contract. After the buyer announces a type θ and the seller chooses a quality q, the message game with the outcomes in Figure 2 is played in stage 3. Then this budget–balanced contract implements the first–best.*

We have already argued why the contract induces the seller to choose the first–best quality and announce message $z_S = T$ in stage 3. By Lemma 2 also the buyer's incentive constraints are satisfied. In particular, condition (29) guarantees that the buyer announces his type truthfully in stage 1. Thus, under the optimal contract the seller can extract the full first–best surplus by choosing k in (30) appropriately.

We may interpret the contract in Proposition 5 as a generalized exit option contract to the extent that the buyer has always the option to exit. In contrast to the more standard exit option contracts from the benchmark cases or Proposition 4, however, the terms of exit also depend on whether the seller agrees to trade or not. This corresponds

[22] If the seller selects the first–best quality, then also (N, N) is an equilibrium of the message game at stage 3. Recall, however, that Condition 1 requires a unique outcome only off the path. Also note that the equilibrium (T, T) Pareto dominates (N, N) at stage 3. Therefore, the two parties can gain from coordinating on (T, T). The multiplicity of equilibria for $q = \tilde{q}(\theta)$ resembles the observation that according to (9) the buyer is indifferent between trade and exit in the benchmark where his type θ is observable.

to an arrangement where both parties have to give their explicit final consent for trade to go ahead. In particular, a court only enforces trade or no trade when statements of both parties are available. Observe also that it is crucial for our contract that the parties submit their messages in ignorance about the other party's message. To facilitate the practical implementation of this, one could employ an impartial mediator who asks each party to confidentially submit their message and then enforces the contract.

7 Conclusion

We have studied bilateral contracting in an environment which is characterized by both contractual incompleteness and asymmetric information. While we derive an impossibility result when trade is deterministic, we explicitly construct first–best efficient mechanisms, with and without third party payments, when stochastic contracts can be enforced. Our mechanisms implement the first–best through simple message games with stochastic terms of trade as off–equilibrium messages which can be interpreted as generalized exit options.

Our insights extend beyond the seller buyer context considered in this paper. Imagine, for example, that the non–verifiable action is more broadly interpreted as some non–contractible decision that an organization has to take. The literature has identified decentralization as a way to mitigate commitment problems that arise from contractual incompleteness. For example, delegating decision rights to lower levels protects subordinates from opportunistic behavior of superiors (e.g. Dessein (2002)). Our efficiency result suggests that centralization is an optimal governance structure once exit options for subordinates are available.

Appendix

Proof of Lemma 1 Let $\theta \in \Theta$. Because the first–best is implementable, the buyer submits the message $m = \theta$ in stage 1 by (19). To simplify notation, we will omit in all functions the dependency on m in the rest of the proof. To prove (25), we begin with two auxiliary steps:

STEP A: We show that for all $\epsilon \in (0, \tilde{q}]$ there is a message $z_B \neq \zeta_B^*(\theta)$ so that

$$v(\tilde{q} - \epsilon, \theta) - p_B^*(\theta) < U(\tilde{q} - \epsilon, \zeta_S^*, z_B \mid \theta). \tag{34}$$

Suppose (34) does not hold. Then there is an $\epsilon \in (0, \tilde{q}]$ such that:

$$v(\tilde{q} - \epsilon, \theta) - p_B^*(\theta) \geq U(\tilde{q} - \epsilon, \zeta_S^*, z_B \mid \theta) \quad \text{for all } z_B \in Z_B. \tag{35}$$

Consider the continuation game $\Gamma(\tilde{q}-\epsilon)$. We show that the message $\zeta_B^*(\theta)$ remains a best response against ζ_S^* in $\Gamma(\tilde{q}-\epsilon)$ for buyer type θ. Indeed, when submitting $\zeta_B^*(\theta)$ in stage 3 in $\Gamma(\tilde{q}-\epsilon)$ against ζ_S^*, trade takes place, and type θ obtains utility $v(\tilde{q}-\epsilon,\theta)-p_B^*(\theta)$. Hence, (35) implies that $\zeta_B^*(\theta)$ is a best response.

Note that by (12), the seller's best response in the message game following the choice of q does not depend on the actual choice of q. Therefore, ζ_S^* remains a best response against $\zeta_B^*(\theta)$ also in the continuation game $\Gamma(\tilde{q}-\epsilon)$. This together with the argument in the previous paragraph proves that if the first–best is implementable, then $(\zeta_S^*, \zeta_B^*(\theta))$ is an equilibrium in $\Gamma(\tilde{q}-\epsilon)$.

Hence, Condition 1 implies that for all equilibria in $\Gamma(\tilde{q}-\epsilon)$, the seller gets the same payment as in $\Gamma(\tilde{q})$. Since the quality $\tilde{q}-\epsilon$ is less costly than $\tilde{q}$, the seller would therefore be better off by choosing quality $\tilde{q}-\epsilon$ at stage 2. This contradicts the condition that in equilibrium $\tilde{q}$ maximizes the seller's profit after receiving the message $m=\theta$, and completes the proof of (34).

STEP B: We next show that (34) implies $x(\zeta_S^*, z_B)=0$. Suppose the contrary. Then, since trade is deterministic, we have $x(\zeta_S^*, z_B)=1$. Thus by (34):

$$v(\tilde{q}-\epsilon,\theta)-p_B^*(\theta) \quad < \quad v(\tilde{q}-\epsilon,\theta)-p_B(\zeta_S^*, z_B). \tag{36}$$

Therefore,

$$\begin{aligned} U(\tilde{q},\zeta_S^*, z_B \mid \theta) \quad &= \quad v(\tilde{q},\theta)-p_B(\zeta_S^*, z_B) && (37)\\ &> \quad v(\tilde{q},\theta)-p_B(\theta,\zeta_S^*,\zeta_B^*(\theta)) = U(\tilde{q},\zeta_S^*,\zeta_B^*(\theta)\mid\theta). && (38)\end{aligned}$$

This means that in the continuation game $\Gamma(\tilde{q})$, type θ would gain by deviating to $z_B \neq \zeta_B^*(\theta)$, a contradiction to the equilibrium condition (13). This completes STEP B.

We can now prove (25). Indeed, by STEP B, $x(\zeta_S^*, z_B)=0$. Thus (34) writes

$$v(\tilde{q}-\epsilon,\theta)-p_B^*(\theta) \quad < \quad -p_B(\zeta_S^*, z_B), \tag{39}$$

establishing (25).

To complete the proof, inequality (26) remains to be shown. But this follows immediately from the equilibrium condition (13) for type θ. Q.E.D.

Proof of Proposition 3 Suppose the first–best is implementable. We derive a contradiction. We begin by showing that the buyer's equilibrium utility

$$U(\theta,\tilde{q}(\theta),\zeta_S^*(\theta),\zeta_B^*(\theta)\mid\theta) \tag{40}$$

is strictly increasing in θ. Indeed, by (15) and (19), buyer type θ does not gain by mimicking the equilibrium strategy of type $\theta' < \theta$. Thus,

$$\begin{aligned} U(\theta, \tilde{q}(\theta), \zeta_S^*(\theta), \zeta_B^*(\theta) \mid \theta) &= v(\tilde{q}(\theta), \theta) - p_B^*(\theta) && (41) \\ &\geq v(\tilde{q}(\theta'), \theta) - p_B(\theta', \zeta_S^*(\theta'), \zeta_B^*(\theta')) && (42) \\ &> v(\tilde{q}(\theta'), \theta') - p_B(\theta', \zeta_S^*(\theta'), \zeta_B^*(\theta')) && (43) \\ &= U(\theta', \tilde{q}(\theta'), \zeta_S^*(\theta'), \zeta_B^*(\theta') \mid \theta'), && (44) \end{aligned}$$

where the strict inequality in the second to last line follows because $v_\theta > 0$ by assumption.

This proves that the buyer's utility is increasing in type θ, and hence there is an $\epsilon > 0$ so that for $\theta > \theta'$:

$$v(\tilde{q}(\theta) - \epsilon, \theta) - p_B^*(\theta) > v(\tilde{q}(\theta'), \theta') - p_B^*(\theta'). \tag{45}$$

Therefore, by Lemma 1 there is a message z_B^E so that

$$-\, p_B(\theta, \zeta_S^*(\theta), z_B^E) > v(\tilde{q}(\theta'), \theta') - p_B^*(\theta'). \tag{46}$$

But observe that the left hand side is the utility that buyer type θ' gets when he submits $m = \theta$ in stage 1 and then chooses z_B^E in stage 3, while the right hand side is the utility that buyer type θ' gets when he tells the truth in stage 1. This contradicts condition (19) which states that first–best implementation requires the buyer to report his type truthfully in stage 1. Q.E.D.

Proof of Lemma 2 We begin by defining $x^E(\theta)$ and $p_B^E(\theta)$. Our general assumptions on v and c imply that $\tilde{q}(\theta)$ is unique and strictly increasing in θ. With the definition of p_B in (30), this implies that

$$v(\tilde{q}(\theta'), \theta') - p_B(\theta') > v(\tilde{q}(\theta), \theta') - p_B(\theta) \quad \text{for all } \theta \neq \theta'. \tag{47}$$

Fix θ. The previous inequality implies that for all $\theta' \neq \theta$, there is an $0 < x(\theta', \theta) < 1$ so that for all $x \in [x(\theta', \theta), 1]$:

$$v(\tilde{q}(\theta'), \theta') - p_B(\theta') \;>\; x[v(\tilde{q}(\theta), \theta') - p_B(\theta)] + (1 - x)[v(\tilde{q}(\theta), \theta) - p_B(\theta)]. \tag{48}$$

Now define

$$\begin{aligned} x^E(\theta) &\equiv \max_{\theta' \neq \theta} x(\theta', \theta), && (49) \\ p_B^E(\theta) &\equiv p_B(\theta) - [1 - x^E(\theta)] v(\tilde{q}(\theta), \theta). && (50) \end{aligned}$$

Because the set of types is finite, we have that $x^E(\theta) < 1$, which in turn implies that $p_B^E(\theta) < p_B(\theta)$, hence (31).

We now show (28). Because $1 - x^E(\theta) > 0$, (50) implies for all $q < \tilde{q}(\theta)$:

$$[1 - x^E(\theta)]v(q, \theta) < p_B(\theta) - p_B^E(\theta), \tag{51}$$

establishing (28).

To show (27) and (29), we first prove that for *all* $\theta', \theta \in \Theta$, it holds:

$$v(\tilde{q}(\theta'), \theta') - p_B(\theta') \geq v(\tilde{q}(\theta), \theta')x^E(\theta) - p_B^E(\theta). \tag{52}$$

Indeed, note that (48) and the definition of $x^E(\theta)$ and $p_B^E(\theta)$ imply for $\theta' \neq \theta$ that

$$\begin{aligned} v(\tilde{q}(\theta'), \theta') - p_B(\theta') &\geq x^E(\theta)[v(\tilde{q}(\theta), \theta') - p_B(\theta)] + (1 - x^E(\theta))[v(\tilde{q}(\theta), \theta) - p_B(\theta)] \\ &= x^E(\theta)v(\tilde{q}(\theta), \theta') - p_B^E(\theta), \end{aligned} \tag{53}$$

establishing (52) for $\theta' \neq \theta$. It is easy to verify that for $\theta' = \theta$, equality holds in (52). From (52), the inequality (27) is direct. Moreover, (29)—with the roles of θ and θ' swapped—now follows from (47) and (52). And this completes the proof. Q.E.D.

Proof of Proposition 4 Lemma 2 and (27)–(28) imply that after reporting honestly in stage 1, the buyer optimally chooses "exit" when the seller selects $q < \tilde{q}(\theta)$ and "trade" when $q \geq \tilde{q}(\theta)$. Therefore, condition (32) implies that the seller optimally selects $\tilde{q}(\theta)$ after receiving message θ in stage 2. Furthermore, by Lemma 2 and (29), the buyer reports his type honestly in stage 1. Finally, observe that Condition 1 holds since after a downward deviation by the seller, the buyer strictly prefers "exit" over "trade" by (28). Q.E.D.

Proof of Proposition 5 We show that there is an equilibrium so that the buyer announces his type honestly in stage 1, the seller selects the first–best quality for this type in stage 2, and trade occurs with probability 1 in stage 3. Moreover, we have to show that Condition 1 is satisfied, that is, when the seller deviates to a quality smaller than first–best, then the equilibrium outcome in stage 3 is unique. We proceed backwards.

STAGE 3: Recall from the main text that when the seller selects quality q, in stage 3 the parties face the normal form game described in Figure 3.

(a) Suppose the buyer was honest in stage 1 and the seller chose $q = \tilde{q}(\theta)$ in stage 2. We show that (T, T) is an equilibrium in stage 3. Indeed, by (27) and because $v(\tilde{q}(\theta), \theta) - c(\tilde{q}(\theta)) > 0$ implies $v(\tilde{q}(\theta), \theta) - p(\theta) > -k + \epsilon$ for ϵ sufficiently small, it follows that $z_B = T$ is a best response by the buyer against $z_S = T$. Moreover, since $p(\theta) > \bar{p}(\theta)$ by (33), $z_S = T$ is a best response by the seller against $z_B = T$. Thus, (T, T) is an equilibrium.

(b) Suppose the buyer was honest in stage 1, but the seller chose $q < \tilde{q}(\theta)$ in stage 2. Then for the buyer $z_B = E$ strictly dominates $z_B = T$ because $v(q,\theta)x^E(\theta) - p^E(\theta) > v(q,\theta) - p(\theta)$ by (28), and $-\max[p^E(\theta),\ k] - \epsilon > -\bar{p}(\theta)$ by (33). After elimination of $z_B = T$, $z_S = N$ strictly dominates $z_S = T$ for the seller because $k > k - \epsilon$ and $\max[p^E(\theta),\ k] + \epsilon > p^E(\theta)$. After elimination of $z_B = T$ and $z_S = T$, $z_B = E$ is strictly dominated for the buyer by $z_B = N$ because $-k > -\max[p^E(\theta),\ k] - \epsilon$. Thus, for $q < \tilde{q}(\theta)$ the game is dominance solvable and the *unique* equilibrium outcome is $(x,p) = (0,k)$. In particular, Condition 1 is satisfied.

STAGE 2: Suppose the buyer was honest in stage 1. We show that the seller optimally chooses the first–best quality $\tilde{q}(\theta)$. Indeed, given the equilibrium outcomes $(1, p(\theta))$ for $q = \tilde{q}(\theta)$ and $(0,k)$ for $q < \tilde{q}(\theta)$ in stage 3, the seller's profit from selecting q is

$$\pi(q) = \begin{cases} k - c(q) & \text{if} \quad q < \tilde{q}(\theta) \\ p(\theta) - c(q) = k + c(\tilde{q}(\theta)) - c(q) & \text{if} \quad q \geq \tilde{q}(\theta) \end{cases} . \tag{54}$$

(To understand the profit expression for $q > \tilde{q}(\theta)$, note that the same arguments as in part (a) imply that if the buyer was honest in stage 1 and the seller selects $q > \tilde{q}(\theta)$, then (T,T) is an equilibrium in stage 3, entailing payments $p(\theta)$.) Hence, by choosing $q = \tilde{q}(\theta)$, the seller's profit is k, while when choosing a different quality, his profit is (weakly) smaller than k. This shows that selecting $q = \tilde{q}(\theta)$ is optimal for the seller in stage 2.

STAGE 1: We show that the buyer announces his type honestly. If type θ reports honestly, his utility is

$$u(\theta) = v(\tilde{q}(\theta),\theta) - p(\theta) = v(\tilde{q}(\theta),\theta) - c(\tilde{q}(\theta)) - k. \tag{55}$$

If he announces $\theta' \neq \theta$, the seller selects $\tilde{q}(\theta')$ in stage 2 and message $z_S = T$ in stage 3. Thus, by selecting an optimal message in stage 3, the buyer obtains

$$\max[-k + \epsilon,\ \ v(\tilde{q}(\theta'),\theta)x^E(\theta') - p^E(\theta'),\ \ v(\tilde{q}(\theta'),\theta) - p(\theta')]. \tag{56}$$

By assumption, we have $v(\tilde{q}(\theta),\theta) - c(\tilde{q}(\theta)) > 0$, hence $u(\theta) \geq -k + \epsilon$ for ϵ sufficiently small. Moreover, by (29), $u(\theta) \geq \max[v(\tilde{q}(\theta'),\theta)x^E(\theta') - p^E(\theta'),\ \ v(\tilde{q}(\theta'),\theta) - p(\theta')]$. Thus, the buyer has no incentive to misreport his type, and this completes the proof. Q.E.D.

8 References

Aghion, P., P. Bolton, and J. Tirole (2004): "Exit Options in Corporate Finance: Liquidity versus Incentives," *Review of Finance*, 8, 327-353.

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

Baliga, S., and T. Sjöström (2009): "Contracting with Third Parties," *American Economic Journal: Microeconomics*, 1, 75-100.

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 T. Chung (1999): "Contract Damages and Cooperative Investments," *RAND Journal of Economics*, 30, 84-105.

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

Che, Y.-K., and J. Sakovics (2004): "A Dynamic Theory of Holdup," *Econometrica*, 72, 1063-1103.

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.

Dessein, W. (2002): "Authority and Communication in Organizations," *Review of Economic Studies*, 69, 811-838.

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.

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

Kahn, C., and G. Huberman (1988): "Two-sided Uncertainty and "Up-or-Out" Contracts," *Journal of Labor Economics*, 6, 423-444.

Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23-38.

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.

Myerson, R. (1986): "Multistage Games with Communication." *Econometrica* 54, 323-58.

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.

Segal, Y. and M. D. Whinston (2002): "The Mirrless Approach to Mechanism Design with Renegotiaton (with Applications to Risk Sharing and the Hold-up Problem)," *Econometrica*, 70, 1-45.

Stiglitz, J. E. and Weiss, A. (1983): "Incentive Effects of Terminations: Applications to the Credit and Labor Markets," *American Economic Review*, 73, 912-927.

Strausz, R. (2006): "Deterministic versus Stochastic Mechanisms in Principal-Agent Models," *Journal of Economic Theory*, 128, 306-314.

Taylor, C. R. (1993): "Delivery-contingent Contracts for Research," *Journal of Law, Economics, and Organization*, 9, 188-203.

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