Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case

Helmut Bester* and Roland Strausz*†

April 1998, revised May 1999,
this revision: May 2000

Abstract

This paper extends the revelation principle to environments in which the mechanism designer cannot fully commit to the outcome induced by the mechanism. We show that he may optimally use a direct mechanism under which truthful revelation is an optimal strategy for the agent. In contrast with the conventional revelation principle, however, the agent may not use this strategy with probability one. Our results apply to contracting problems between a principal and a single agent. By reducing such problems to well–defined programming problems they provide a basic tool for studying imperfect commitment.

Keywords: revelation principle, mechanism design, contract theory, limited commitment, asymmetric information; *JEL Classification No.: D82, C72*

*Free University Berlin, Dept. of Economics, Boltzmannstr. 20, D-14195 Berlin (Germany); Email addresses: *hbester@wiwiss.fu-berlin.de* and *strausz@wiwiss.fu-berlin.de*, respectively.

†We wish to thank three referees, Drew Fudenberg and Klaus Schmidt for helpful comments. We are grateful to Hans Haller and Andras Löffler for their mathematical assistance.

1 Introduction

This paper provides a modified version of the revelation principle for environments in which the party in the role of the mechanism designer cannot fully commit to the outcome induced by the mechanism. This version is a prerequisite for solving contracting problems between a principal and a single agent in situations where contractual commitments are limited. The revelation principle, which has been shown by Gibbard (1973), Green and Laffont (1977), Dasgupta *et. al.* (1979) and Myerson (1979), is the guiding principle for the theory of implementation and mechanism design under imperfect information. It states that the range of implementable outcomes is simply the set of outcomes that give no incentive to the agent to misrepresent his type. This effectively reduces the problem of finding an optimal mechanism within the set of all conceivable mechanisms to a straightforward maximization problem subject to so–called incentive compatibility constraints. This reduction of complexity makes the revelation principle the universal starting point for the analysis of contracting and mechanism design problems.

Yet, an important drawback of the revelation principle is that it is only applicable to settings in which the mechanism designer is able to credibly commit to any outcome of the mechanism.[1] Especially for contractual relationships this requirement has implications that may be unrealistic. First, in a long–term relationship the mechanism designer has to specify a contract that covers the entire time horizon of the relationship. Second, he must be able to resist renegotiating away ex post inefficiencies. This is a serious problem because under asymmetric information the ex ante optimal contract typically exhibits ex post inefficient outcomes. To assume that such contracts will not be renegotiated somehow runs against the central postulate of contract theory that all benefits from trade will be exploited. Third, any action that the mechanism designer may take must be verifiable so that it can be specified as part of the mechanism.

If any of these conditions is not met, the argument of the conventional revelation principle fails for a simple reason: Suppose that, in accordance with truthful revelation, the agent reveals his type. Then the mechanism designer is practically fully informed. Clearly, he will exploit this information for all those decisions to which he is not committed by the mechanism. Since

[1] The revelation principle may fail not only because of imperfect commitment; it also does not apply to situations where several mechanism designers compete against each other. We do not address the second case, which is studied by Epstein and Peters (1996) and Martimort and Stole (1999).

the agent anticipates this, he may realize that truthfully reporting his private information is disadvantageous for him. Therefore, the standard revelation principle fails with imperfect commitment. In fact, in settings of imperfect commitment it is easy to construct mechanisms whose outcome cannot be replicated by a direct mechanism. This observation makes the analysis of implementable allocations highly complicated as there is no obvious restriction on the set of conceivable mechanisms.

In a contracting problem, however, the parties care about implementable allocations and the communication game only insofar as their *payoffs* are affected. This argument, which is used also in Maskin and Tirole (1999), allows us to extend the two most essential features of the revelation principle to situations with imperfect commitment. Indeed, we are able to prove that the payoffs on the Pareto frontier of an arbitrary mechanism may also be obtained by a direct mechanism, in which the agent's message space is simply the set of his types. Further, under this mechanism it is an optimal strategy for the agent to reveal his type truthfully and he will use this strategy with positive probability. As an important consequence, also in the presence of imperfect commitment an optimal mechanism can still be found in the set of incentive compatible direct mechanisms. In contrast with the full commitment case, however, incentive compatibility of truthful reporting is no longer sufficient for implementation. This is so because typically the agent has to be kept indifferent between truthfully revealing his information and cheating, which may occur with positive probability. Nonetheless, our version of the revelation principle allows us to formulate the mechanism designer's problem as a straightforward programming problem. We thus provide a basic tool for studying mechanism design problems that involve sequential contracting, renegotiation or incomplete contracts.

Our result is relevant for the large part of contract theory that studies contracting problems between a principal and a single agent. In this area, the analysis of limited commitment has become a major direction of research. This research is generally well aware of the unfortunate gap in implementation theory that arises because the derivation of optimal contracts cannot appeal to the conventional revelation principle. Filling this gap is the main purpose of this paper.

In many cases, the literature on contracting with limited commitment simply sidesteps this problem by imposing artificial restrictions on the form of contracts. This approach is not fully satisfactory as the 'optimal' contract under these restrictions may not be fully optimal within the set of all conceivable contracts. In Freixas, Guesnerie, and Tirole (1988), for instance,

the principal is constrained to linear contracts and cannot use a revelation mechanism. Similarly, Dewatripont's (1989) analysis of renegotiation–proof labor contracts simply imposes incentive compatibility restrictions without justifying them. A justification for this procedure is given in this paper.[2] In their study of regulatory dynamics without commitment, Laffont and Tirole (1993, chapter 9) focus on the two–type case and restrict the regulator to offering a menu of two contracts. They acknowledge that they "did not find any argument to exclude generally menus with more than two contracts" (p. 390). This paper provides an argument for why their approach involves no loss of generality.

Only a small part of the literature explicitly allows for general mechanisms to derive optimal contracts (Hart and Tirole (1988), Laffont and Tirole (1990)). These papers, however, do not present a generalizable concept to characterize the structure of optimal contracts. Rather, they derive for a particular setting some specific form of the general result established here.

The rest of the paper is organized as follows. Section 2 provides a general model of contracting with imperfect commitment. It also contains an example which we use in the following sections to illustrate our results. In Section 3 we show first that with a message set of the same dimensionality as the set of the agent's types, the mechanism designer can get the same payoff as from solving a contracting problem with an arbitrary message set. In Section 4, we prove in a second step the optimality of a direct mechanism under which the agent has a weak incentive to reveal his type truthfully. The application of our results enables us to state the contracting problem with imperfect commitment as a well–defined optimization problem, which we describe in Section 5. Section 6 shows that multi–stage mechanism design problems can be reduced to standard dynamic programming problems. All proofs are relegated to an Appendix.

2 Contracting with Limited Commitment

We consider a contracting problem between a principal and a single agent in an adverse selection environment. The solution of this problem determines an allocation $z = (x, y) \in Z = X \times Y$. An allocation consists of two types of decision variables: By X we denote all those decisions to which the principal can contractually commit himself. In contrast, Y describes all those decisions

[2] At the same time, however, Dewatripont (1989) does not allow the informed party to randomize. Given our results, this seems restrictive.

that are not contractible and are chosen at the principal's discretion. Both X and Y are taken to be metric spaces and we denote by $\mathcal{X}$ and $\mathcal{Y}$ the Borel $\sigma-$algebra on X and Y, respectively. We allow for the possibility that the decision x restricts the principal's feasible choices in Y and describe this restriction by a correspondence $F: X \Rightarrow Y$.[3] Thus, the principal has to select $y \in F(x)$ when he is committed to the decision $x \in X$.

The principal has no private information. The agent, however, is privately informed about his type $t \in T = \{t_1, ..., t_i, ..., t_{|T|}\}$. We assume that $2 \leq |T| < \infty$.[4] The principal only knows the probability distribution $\gamma = (\gamma_1, ..., \gamma_i, ..., \gamma_{|T|})$ of the agent's type, with $\gamma_i > 0, i = 1, ..., |T|$, and $\sum_i \gamma_i = 1$. The payoffs of both players depend on the allocation (x, y) and the agent's type: When the agent is of type t_i, the principal's payoff from (x, y) is $V_i(x, y)$. The agent's payoff in this situation is $U_i(x, y)$. Both $V_i(\cdot)$ and $U_i(\cdot)$ are assumed to be continuous and bounded on Z.

To solve his problem, the principal will require the agent to provide some kind of information. He therefore chooses a message set M so that the agent has to select some message $m \in M$. Let M be a metric space and let $\mathcal{M}$ denote the Borel $\sigma-$algebra on M. The principal can commit himself to a measurable decision function $x: M \to X$. The interpretation is that, once the principal has committed himself to the decision function $x(\cdot)$, the agent can enforce the decision $x(m)$ by sending the message m. A *mechanism* or contract $\Gamma = (M, x)$ specifies a message set M in combination with a decision function $x(\cdot)$.

A mechanism Γ induces the following game between the principal and the agent: First, the agent selects some message $m \in M$. This determines the contractually specified decision $x(m) \in X$. The principal uses the agent's message to update his beliefs about the agent's type and chooses some decision $y \in F(x(m))$. Given Γ, the principal is constrained to the allocations that can be obtained through the Perfect Bayesian Equilibria of this game.

More formally, the agent's strategy in this game is a mapping $q: T \to Q$, where Q is the set of probability measures on $\mathcal{M}$. Thus, if $q_i(H) > 0$ for some $t_i \in T$ and $H \in \mathcal{M}$, this means the message chosen by the t_i-type agent lies in H with probability $q_i(H)$. Let $\bar{q} \equiv \sum_i \gamma_i q_i$ and note that $\bar{q} \in Q$. The principal's strategy is a measurable function $y: M \to F(x(m))$. Thus $y(m)$ describes his choice as a function of the observed message m. We denote

[3]For every $x \in X$, the $\sigma-$algebra on $F(x)$ is taken to be the Borel $\sigma-$algebra which is naturally induced by $\mathcal{Y}$.

[4]Throughout we denote by $|A|$ the cardinality of a set A.

the principal's posterior belief by the measurable mapping $p\colon M \to P$, where $P = \{p \in \mathbb{R}_+^{|T|} \,|\, \sum_i p_i = 1\}$ is the set of probability distributions over T. Thus, when the principal observes the message $m \in M$, he believes that the agent is of type t_i with probability $p_i(m)$. For a given contract $\Gamma = (M, x)$ and a given strategy combination (q, y), the expected payoffs for the principal and the t_i-type agent are defined as

$$\begin{aligned} V^*(q, y, x|M) &= \sum\nolimits_i \gamma_i \int_M V_i(x(m), y(m)) \mathrm{d}q_i(m), \qquad (1) \\ U_i^*(q, y, x|M) &= \int_M U_i(x(m), y(m)) \mathrm{d}q_i(m). \end{aligned}$$

To constitute a Perfect Bayesian Equilibrium, the functions (q, p, y) have to satisfy three conditions: First, the principal's strategy has to be optimal given his beliefs about the agent's type. This means that, for every $m \in M$,

$$\sum\nolimits_i p_i(m)\, V_i(x(m), y(m)) \geq \sum\nolimits_i p_i(m)\, V_i(x(m), y') \quad \text{for all } y' \in F(x(m)). \quad (2)$$

Second, the agent anticipates the principal's behavior y and chooses q to maximize his payoff. Thus, for each $t_i \in T$, q_i has to satisfy

$$\int_M U_i(x(m), y(m)) \mathrm{d}q_i(m) \geq \int_M U_i(x(m), y(m)) \mathrm{d}q_i'(m) \quad \text{for all } \; q_i' \in Q. \quad (3)$$

Finally, the principal's posterior belief has to be consistent with Bayes' rule on the support of the agent's message strategy, except possibly for a set of messages that have measure zero under this strategy. That is, for all $t_i \in T$ and all $H \in \mathcal{M}$ with $\bar{q}(H) > 0$ it is required that

$$\int_H p_i(m) \mathrm{d}\bar{q}(m) = \gamma_i q_i(H). \quad (4)$$

By this condition, the principal's posterior is compatible with Bayesian updating: After dividing both sides of (4) by $\bar{q}(H) > 0$, the left hand side represents the principal's belief that he confronts agent t_i upon receiving a message from the set H. The right hand side expresses the conditional probability that the agent is actually of type t_i given that, under the reporting strategy q, a message in the set H is realized. By Radon–Nikodym's Theorem (see e.g. Stokey and Lucas (1989), p. 194), equation (4) defines p uniquely $\bar{q}-$almost everywhere.[5]

[5] In terms of the Radon–Nikodym derivative, condition (4) may be expressed as $p_i(m) = [\mathrm{d}\gamma_i q_i(m)]/[\mathrm{d}\bar{q}(m)]$. Of course, the belief p determines the principal's choice of y also for out–of–equilibrium messages. Yet, Bayes' rule imposes no restrictions on beliefs outside the support of the agent's reporting strategy,

We say that $(q, p, y, x|M)$ is *incentive feasible* if (q, p, y) is a Perfect Bayesian equilibrium given the mechanism Γ. For a given message set M, $(q, p, y, x|M)$ is said to be *incentive efficient* if it is incentive feasible and there is no incentive feasible $(q', p', y', x'|M)$ such that

$$V^*(q', y', x'|M) > V^*(q, y, x|M) \text{ and } U_i^*(q', y', x'|M) = U_i^*(q, y, x|M) \quad (5)$$

for all $t_i \in T$.[6] Finally, $(q, p, y, x|M)$ and $(q', p', y', x'|M')$ are *payoff–equivalent* if $V^*(q, y, x|M) = V^*(q', y', x'|M')$ and $U_i^*(q, y, x|M) = U_i^*(q', y', x'|M')$ for all $t_i \in T$.

In most contracting problems, the agent has the option to refuse to contract with the principal. If we let $\bar{U}_i$ denote the payoff that the type t_i-agent can get without cooperating, then $(q, p, y, x|M)$ must also satisfy the individual–rationality constraints

$$U_i^*(q, y, x|M) \geq \bar{U}_i \quad \text{for all} \quad t_i \in T. \quad (6)$$

These constraints guarantee that the agent can obtain his reservation payoff *ex post* after he has learned his type.[7] Indeed, (6) ensures that for each type t_i there is a message $m \in M$ such that $U_i(x(m), y(m)) \geq \bar{U}_i$. When the contract is concluded before the agent knows his type, the principal only has to observe the *ex ante* constraint $\sum_i \gamma_i U_i^* \geq \sum_i \gamma_i \bar{U}_i$ instead of the stronger condition (6). Although we focus our discussion on the case where (6) is relevant, it should be clear that our results also cover problems in which the contract only has to satisfy an ex ante individual–rationality constraint for the agent.

For a given M, the principal's problem is to find a (q, p, y, x) that maximizes his expected payoff subject to the constraints (2)–(4) and (6). Obviously, any solution to this problem must be incentive efficient. In addition, of course, the principal's overall problem includes the choice of an appropriate message set M. In the following sections we illustrate our analysis of this problem by an example which is adopted from Miyazaki (1977).

Example There are two types of the agent; each type is equally likely. The principal chooses the agent's speed s of work and pays the wage w. The

[6] When the set of incentive feasible outcomes is non–empty, one may guarantee the existence of an incentive efficient outcome by assuming that M and Z are compact and that $F(\cdot)$ is continuous.

[7] For convenience, we assume throughout that the principal wants all types to participate, which is optimal if his outside option from not employing a particular type is small enough. Our results, however, are applicable also in the case where the inequality in (6) is reversed for some types.

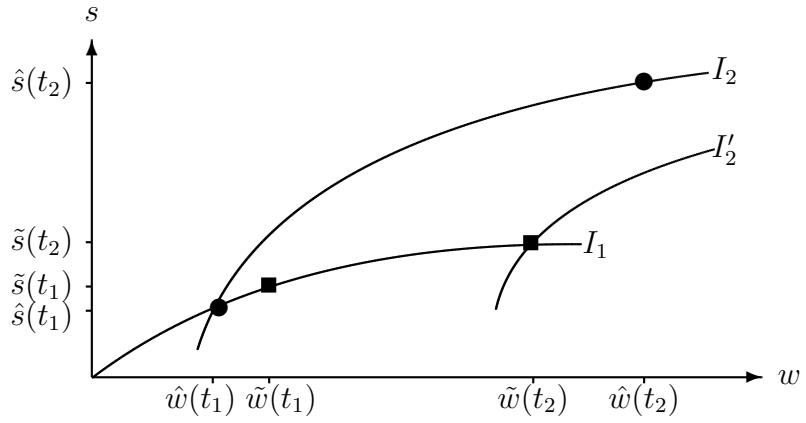

Figure 1: *The Agent's Indifference Curves*

principal's and the agent's payoffs are

$$V_1(w,s) = 10s - s^2 - w, \quad V_2(w,s) = 10s - s^2/4 - w, \tag{7}$$

$$U_1(w,s) = w - s^2/5, \quad U_2(w,s) = w - s^2/6, \quad \bar{U}_1 = \bar{U}_2 = 0.$$

Notice that the agent's utility functions satisfy the 'single–crossing property'. This is illustrated in Figure 1 which depicts two indifference curves, I_2 and I_2', for the type t_2 agent and the indifference curve I_1, along which the payoff of type t_1 equals zero. When both w and s are contractible, the principal can fully commit himself and $Z = X = \{(w,s)|w \geq 0, s \geq 0\}$. If however s is not verifiable and hence not contractible, the principal can commit himself only to w. In this case $X = \{w|w \geq 0\}$ and $Y = \{s|s \geq 0\}$. As a reference point, we note that the first best solution under perfect information satisfies $s^*(t_i) = \text{argmax}\,(V_i + U_i)$, which yields $s^*(t_1) = 25/6$ and $s^*(t_2) = 12$. $\diamond$

To emphasize the generality of our framework, we conclude this section by some further examples of the environment described above.

Unobservable actions As in our example, the principal's decision y may not be contractible because it is not publicly observable. Another example is the model of Khalil (1997), where the principal can perform an audit to verify the agent's type, but commitment to a specific auditing strategy is not possible.

Short–term contracts As in Freixas et al. (1985) and Laffont and Tirole (1987), imperfect commitment arises in a two-period contracting model in which contracts binds the contracting parties for the current period only. That is, after the contract x expires, the principal offers a new contract y for the second period.

Renegotiation When the principal receives the message m, he may realize that the contract $x^*(m)$ is ex–post inefficient. As in Laffont and Tirole (1990) he will then propose a new contract y, which the agent can either accept or reject. Since the agent can insist on the original contract $x^*(m)$, this imposes a restriction $F(x^*(m))$ on the set of contracts Y that the principal can achieve through renegotiation.

Cheap talk When $Z = Y$, the principal faces no commitment and so the agent's message m has no direct impact on the allocation. It can only play a role when it affects the principal's decision via his beliefs about the agent's type. In this situation the revelation game reduces to a cheap talk game as in Crawford and Sobel (1982).

As we show in Section 6, one can apply our approach also to multi–stage mechanism design problems. By the technique of dynamic programming, such a problem can be stated as a sequence of static problems. This allows us to treat each contracting stage in the terms of our model.

3 Direct Mechanisms

The mechanism $\Gamma = (M, x)$ is said to be a *direct* mechanism if $M = T$. With a direct mechanism the message set is the agent's type set and in the game induced by Γ the agent simply announces some type.

As a special case of the environment described in the previous section, the principal may be able to commit himself to all relevant decisions. This is the case when $Z = X$. Then conditions (2) and (4) are redundant and condition (3) defines an optimal reporting strategy for the agent. In this situation, it is well–known that the following important observation applies:[8]

Revelation Principle *Assume $Z = X$. If $(q, x|M)$ is incentive feasible, then there exists a direct mechanism $\Gamma^d = (T, \hat{x})$ and an incentive feasible $(\hat{q}, \hat{x}|T)$ such that $(\hat{q}, \hat{x}|T)$ and $(q, x|M)$ are payoff–equivalent. Moreover, $\hat{q}_i(t_i) = 1$ for all $t_i \in T$.*

The revelation principle greatly simplifies the principal's problem for two reasons: First, it allows him to select as his message set simply the set of the agent's types. Second, he can restrict himself to an allocation function

[8]The standard revelation principle requires that the principal can commit to probabilistic decisions. Therefore, the following statement assumes that X represents the set of probability distributions over deterministic decisions and that $U_i(x)$ and $V_i(x)$ are the expected payoffs from the distribution x.

$z: T \to X$ which gives the agent no incentive to misrepresent his type. Thus, for each $t_i \in T$, we can replace the constraint (3) by the *incentive compatibility* conditions

$$U_i(z(t_i)) \geq U_i(z(t_j)) \quad \text{for all} \quad t_j \in T. \tag{8}$$

The incentive compatibility conditions effectively summarize the restrictions that the principal faces because he is uninformed about the agent's type. For an incentive compatible allocation, the individual rationality constraints (6) become

$$U_i(z(t_i)) \geq \bar{U}_i, \quad \text{for all} \quad t_i \in T. \tag{9}$$

Thus, in the case of full commitment, the revelation principle reduces the contracting problem to a straightforward programming problem: The principal has to select an allocation function $z(\cdot)$ that maximizes his expected payoff $\sum_i \gamma_i V_i(z(t_i))$ subject to the incentive compatibility and individual rationality constraints (8) and (9). We now derive the solution of this problem for our example in Section 2.

EXAMPLE With full commitment, the principal chooses $z(t_1) = (w(t_1), s(t_1))$ and $z(t_2) = (w(t_2), s(t_2))$ to maximize his expected payoff subject to the constraints (8) and (9). Since $\bar{U}_1 = \bar{U}_2 = 0$, it is easy to show that because of the single–crossing property only the constraints

$$U_2(w(t_2), s(t_2)) = U_2(w(t_1), s(t_1)), \quad U_1(w(t_1), s(t_1)) = 0, \tag{10}$$

are binding. Maximizing $0.5[V_1(w(t_1), s(t_1)) + V_2(w(t_2), s(t_2))]$ subject to (10) yields the allocation function

$$\hat{s}(t_1) = 150/37,\ \hat{s}(t_2) = 12,\ \hat{w}(t_1) = 4500/1369,\ \hat{w}(t_2) = 33606/1369, \tag{11}$$

which is illustrated in Figure 1. A comparison with the first best solution under perfect information shows that $\hat{s}(t_1) < s^*(t_1)$ and $\hat{s}(t_2) = s^*(t_2)$. Thus there is 'no distortion at the top'. Also, the type t_2 agent earns an 'informational rent' because $U_2(\hat{w}(t_2), \hat{s}(t_2)) > 0$. $\diamond$

The underlying idea of the standard revelation principle is to combine the two functions $q: T \to Q$ and $x: M \to X$ to a single function $\hat{x}: T \to X$. In this way, any incentive feasible $(q, x|M)$ can be replaced by an incentive feasible $(\hat{q}, \hat{x}|T)$ that induces the same probability distribution over allocations. At least this is possible as long as the principal can commit himself to the entire allocation $\hat{z} = \hat{x}$. With imperfect commitment, however, the procedure of the

standard revelation principle is no longer applicable. We demonstrate this in the setting of our example by showing that a contract Γ may support an outcome that cannot be replicated by a direct mechanism Γ^d.

EXAMPLE Suppose the principal can commit himself to the wage w but not to the speed of work s so that $X = \{w|w \geq 0\}$ and $Y = \{s|s \geq 0\}$. Consider the message set $M = \{m_1, m_2, m_3\}$ and let

$$s(m_1) = 5, \quad s(m_2) = 10, \quad s(m_3) = 20, \tag{12}$$

$$w(m_1) = 5, \quad w(m_2) = 20, \quad w(m_3) = 70.$$

Since $s = 20/(1 + 3\,p_1)$ maximizes $p_1 V_1(w, s) + (1 - p_1)V_2(w, s)$, it follows that the belief

$$p_1(m_1) = 1, \quad p_1(m_2) = 1/3, \quad p_1(m_3) = 0. \tag{13}$$

supports the principal's choice of $s(\cdot)$. Therefore, (12) and (13) satisfy condition (2). Condition (3) is satisfied for the agent's reporting strategy

$$q_1(m_1) = 3/4, \quad q_1(m_2) = 1/4, \quad q_1(m_3) = 0; \tag{14}$$

$$q_2(m_1) = 0, \quad q_2(m_2) = 1/2, \quad q_2(m_3) = 1/2.$$

because by (7) and (12)

$$\begin{aligned} U_1(w(m_1), s(m_1)) = \;& U_1(w(m_2), s(m_2)) > \; U_1(w(m_3), s(m_3)), \\ U_2(w(m_2), s(m_2)) = \;& U_2(w(m_3), s(m_3)) > \; U_2(w(m_1), s(m_1)). \end{aligned} \tag{15}$$

The belief in (13) is consistent with (14) and so condition (4) is satisfied. Thus $(q, p, s, w|M)$ is incentive feasible. Since $q_1(m_h) + q_2(m_h) > 0$ for all $m_h \in M$, each $s(m_h)$ is implemented with positive probability. With the message set $T = \{t_1, t_2\}$, however, condition (2) implies that at most two different values $s(t_1)$ and $s(t_2)$ can be implemented because the principal's payoff is strictly concave in s. $\diamond$

At least at first sight, the design of efficient mechanisms or optimal contracts faces a serious problem from the observation that a direct mechanism may not support all outcomes that are implementable through some other type of mechanism. It is therefore unclear how to characterize the set of implementable allocations. Yet, as we argued in the Introduction, we can circumvent this problem by focusing on payoffs rather than allocations. As we demonstrate in the remainder of this section, any incentive efficient $(q, p, y, x|M)$ can be replaced by some payoff–equivalent $(q', p, y, x|M)$ in such

a way that the support of the agent's reporting strategy q' contains at most $|T|$ elements. This means that the principal can solve his contracting problem by using a message set of dimension $|T|$.

To prove the main result of this section, we employ the following technical lemma, which states a first–order condition implied by incentive efficiency.

Lemma 1 *Let $(q, p, y, x|M)$ be incentive efficient. Then there exists a $\mu = (\mu_1, ..., \mu_i, ..., \mu_{|T|}) \in \mathbf{R}^{|T|}$ such that*

$$\sum\nolimits_i V_i(x(m), y(m)) p_i(m) = \sum\nolimits_i \frac{\mu_i}{\gamma_i} p_i(m) \quad \bar{q} - \textit{almost everywhere.}$$

We now use Caratheodory's theorem (see e.g. Rockafellar (1970), p. 155) to derive from q another reporting strategy q' that also satisfies the efficiency condition of Lemma 1. This leads us to the following result.

Proposition 1 *Let $(q, p, y, x|M)$ be incentive efficient. Then there exist an incentive feasible $(q', p, y, x|M)$ and a finite set $M' = \{m_1, ..., m_h, ...m_{|M'|}\} \in \mathcal{M}$ with $|M'| \leq |T|$ and $\bar{q}\,'(M') = 1$ such that $(q, p, y, x|M)$ and $(q', p, y, x|M)$ are payoff–equivalent. Moreover, the vectors $q'(m_h) = (q'_1(m_h), ..., q'_i(m_h), ..., q'_{|T|}(m_h))$, $h = 1, ..., |M'|$, are linearly independent.*

The basic idea for deriving the reporting strategy q' can easily be explained for the case where $M = \{m_1, ..., m_h, ..., m_{|M|}\}$ is a finite set. If a mechanism uses more messages than types, then the vectors $q(m_h) \neq 0$, $h = 1, ..., |M|$, are necessarily linearly dependent. This means that some messages are redundant: The principal's belief p can be made consistent with Bayes' rule in (4) also if the agent selects a reporting strategy q' that uses at most $|T|$ messages. To see this, suppose for simplicity that $\bar{q}(m_h) > 0$ for all $m_h \in M$. Since Bayes' rule then implies $\sum_h p_i(m_h)\bar{q}(m_h) = \gamma_i$, the vector γ lies in the convex hull of the vectors $p(m_h) = (p_1(m_h), ..., p_{|T|}(m_h))$, $h = 1, ..., |M|$. The dimension of this convex hull is smaller or equal to $|T| - 1$ because $\sum_i p_i(m_h) = 1$. Caratheodory's theorem therefore asserts that

$$\sum\nolimits_h \alpha_h\, p(m_h) = \gamma \tag{16}$$

has a non–negative solution α such that at most $|T|$ of the scalars $\alpha_h, h = 1, ..., |M|$, are positive. This allows us to define the reporting strategy q' by setting $q'_i(m_h) \equiv \alpha_h\, q_i(m_h)/\bar{q}(m_h)$ because (4) and (16) guarantee that $\sum_h q'_i(m_h) = \sum_h \alpha_h\, p_i(m_h)/\gamma_i = 1$. Obviously, the support of $\bar{q}\,'$ contains at

most $|T|$ messages and p and q' are consistent with Bayesian updating. Note that the support of $\bar{q}'$ is contained in the support of $\bar{q}$. This means that the allocations which are realized under $(q', p, y, x|M)$ are a subset of the allocations that can occur under $(q, p, y, x|M)$.

Since replacing q by q' does not alter the principal's belief, his choice of y remains optimal. Also, using the fact that q' satisfies the efficiency condition stated in Lemma 1, one can show that the principal's expected payoff is not changed. Finally, the agent's reporting behavior q' is optimal and he receives the same expected payoff under q and q', because he is indifferent between all messages that he selects with positive probability. The latter point also explains our observation in Bester and Strausz (2000) that Proposition 1 may fail in environments with multiple agents. The reason is that applying the above procedure to the reporting behavior of one of the agents may reduce the expected payoff of *another* agent so that individual rationality may be violated.

4 Incentive Compatibility

Proposition 1 reduces the complexity of the contracting problem drastically. It allows the principal to disregard message sets which contain more messages than the agent's types. Yet, Proposition 1 is not very specific about the message set and the agent's reporting behavior. Its second part, however, provides information that we can exploit to reduce the complexity of the contracting problem even further. In addition, it enables us to formulate our results in such a way that we can relate them to incentive compatibility conditions implied by the standard revelation principle.

Our argument uses the linear independence of the vectors $q'(m_h)$, $h = 1, ..., |M'|$, in Proposition 1 to apply the classical marriage theorem. This theorem, which we restate in the next lemma, was first stated and shown in its definitive form in Hall (1935) and Maak (1936). Weyl (1949) introduced the term 'marriage theorem'.[9] We rely on a statement and a proof found in Jacobs (1969, pp. 105–106).

[9]The interpretation of Lemma 2 as the marriage theorem identifies H with a set of men, where each man h has some acquaintances $D(h)$ in the set of women K. The mapping d represents a marriage which by (i) respects monogamy and by (ii) associates each man only with one of his acquaintances. We are grateful to Hans Haller for informing us about the marriage theorem and its usefulness in the present context.

Lemma 2 *Let H be a finite non–empty set and K be a non–empty set, possibly infinite. Further, let $D\colon H \Rightarrow K$ be a correspondence and for any set $G \subseteq H$ define*

$$D(G) = \bigcup_{h \in G} D(h).$$

Then there exists a mapping $d\colon H \to K$ such that
(i) $d(h) = d(k)$ implies $h = k$, and
(ii) $d(h) \in D(h)$ for all $h \in H$,
if and only if $|D(G)| \geq |G|$ for all $G \subseteq H$.

We apply the marriage theorem to the sets M' and T in our Proposition 1 by defining $D\colon M' \Rightarrow T$ so that $D(m_h)$ denotes the set of all types that use message m_h with positive probability. By the theorem we can then associate with each message $m_h \in M'$ a type $d(m_h) \in D(m_h)$ in such a way that two distinct messages are never associated with the same type. This allows us to construct a direct mechanism whose properties are very similar to the standard revelation principle.

Proposition 2 *If $(q, p, y, x|M)$ is incentive efficient, then there exists a direct mechanism $\Gamma^d = (T, \hat{x})$ and an incentive feasible $(\hat{q}, \hat{p}, \hat{y}, \hat{x}|T)$ such that $(\hat{q}, \hat{p}, \hat{y}, \hat{x}|T)$ and $(q, p, y, x|M)$ are payoff–equivalent. Moreover, $\hat{q}_i(t_i) > 0$ for all $t_i \in T$.*

The differences between Proposition 2 and the conventional revelation principle indicate the implications of imperfect commitment: While the conventional revelation principle applies whenever $(q, p, y, x|M)$ is incentive feasible, our result uses the stronger requirement of incentive efficiency. Yet, as we argued above, this does not restrict the usefulness of our result as a tool for the design of efficient contracts. The other difference between Proposition 2 and the conventional revelation principle is that we can no longer guarantee that the agent reveals his type with certainty. Still, as $\hat{q}_i(t_i) > 0$, truthful reporting is an optimal strategy for the agent and he chooses this strategy with positive probability.[10]

5 Optimal Contracts

The important insight from Proposition 2 is that with a direct mechanism the principal can get at least the same payoff as from solving his contracting

[10]This does not preclude that two types t_i and t_j choose a fully pooling strategy with $q_i(t) = q_j(t)$ for all $t \in T$.

problem with some other message set. Moreover, without loss of generality he may restrict himself to an incentive compatible allocation function $z = (x, y)\colon T \to Z$, under which no type t_i of the agent can gain by reporting some other type $t_j \neq t_i$. These two observations allow us to formulate the principal's contracting problem as a straightforward programming problem:

$$\begin{aligned}
\max_{q,p,y,x} \quad & \textstyle\sum_i \sum_j \gamma_i q_i(t_j) V_i(z(t_j)) && (17)\\
\text{s.t.} \quad & U_i(z(t_i)) \geq U_i(z(t_j)), && (18)\\
& U_i(z(t_i)) \geq \bar{U}_i, && (19)\\
& [U_i(z(t_i)) - U_i(z(t_j))]\; q_i(t_j) = 0, && (20)\\
& y(t_i) \in \operatorname*{argmax}_{y \in F(x(t_i))} \textstyle\sum_j p_j(t_i)\, V_j(x(t_i), y), && (21)\\
& p_i(t_j) \textstyle\sum_k \gamma_k\, q_k(t_j) = \gamma_i\, q_i(t_j), && (22)
\end{aligned}$$

for all $t_i, t_j \in T$. The first two constraints represent the usual incentive compatibility and individual rationality restrictions. Because of his inability to commit himself to a decision $y \in Y$, the principal faces three additional constraints. First, whenever it is optimal to induce the t_i-type agent to report some $t_j \neq t_i$, he has to be kept indifferent between reporting t_i and t_j, which is equivalent to condition (20). Second, the principal's choice of y has to satisfy condition (2) so that (21) must be satisfied.[11] Finally, the condition of Bayesian consistency in (4) requires (22). We now apply this programming problem to derive the solution for our example.

Example When the principal cannot commit himself to s, constraint (21) requires his choice of $s(t_i)$ to maximize $p_1(t_i)V_1(w, s) + (1 - p_1(t_i))V_2(w, s)$. This implies

$$s(t_i) = 20/\left(1 + 3\, p_1(t_i)\right). \tag{23}$$

It is easy to see that because of the single–crossing property the incentive compatibility constraint is binding only for one type of the agent. Consider the case where this is the agent of type t_1. Then we have $\tilde{q}_2(t_1) = 0, \tilde{q}_2(t_2) = 1$ and

$$U_1(w(t_2), s(t_2)) = U_1(w(t_1), s(t_1)), \quad U_1(w(t_1), s(t_1)) = 0, \tag{24}$$

and this is sufficient to guarantee that (18), (19) and (20) are satisfied. By (22) the principal's beliefs are

$$p_1(t_1) = 1, \quad p_1(t_2) = [1 - q_1(t_1)]/[2 - q_1(t_1)]. \tag{25}$$

[11] In some applications this condition can be more conveniently formulated as a first–order condition.

He thus maximizes $0.5[q_1(t_1)V_1(w(t_1), s(t_1)) + (1 - q_1(t_1))V_1(w(t_2), s(t_2)) + V_2(w(t_2), s(t_2))]$ subject to (24) and

$$s(t_1) = 5, \quad s(t_2) = [40 - 20q_1(t_1)]/[5 - 4q_1(t_1)]. \tag{26}$$

The solution yields $\tilde{q}_1(t_1) = \tilde{q}_1(t_2) = 1/2$ and the allocation function[12]

$$\tilde{s}(t_1) = 5, \ \tilde{s}(t_2) = 10, \ \tilde{w}(t_1) = 5, \ \tilde{w}(t_2) = 20, \tag{27}$$

which is illustrated in Figure 1. The principal's expected payoff is $55/2$. This is indeed the highest payoff that he can get because by the same procedure one may verify that by keeping the type t_2 agent indifferent between $(w(t_1), s(t_1))$ and $(w(t_2), s(t_2))$ he can only get $136/5$. Interestingly, the comparison with the first–best outcome s^* and the full commitment solution $\hat{s}$ shows that $\tilde{s}(t_1) > s^*(t_1) > \hat{s}(t_1)$ and $\tilde{s}(t_2) < s^*(t_2) = \hat{s}(t_1)$. $\diamond$

As the example illustrates, our analysis simplifies the task of finding an optimal contract to a mere computational problem. Typically, the main technical difficulty in deriving the solution is to find out which of the incentive constraints are binding at the optimum. This difficulty occurs already in mechanism design problems with perfect commitment. In many applications regularity conditions such as the single–crossing property are helpful to identify binding incentive constraints. With imperfect commitment, however, the single–crossing condition is in general no longer sufficient to detect the relevant incentive restrictions easily.[13] As Laffont and Tirole (1993, p. 377) note, due to the lack of commitment "any incentive constraint could turn out to be binding at the optimum". Therefore, one may have to resort to the straightforward but tedious procedure of solving the optimization problem by examining all possible constellations case by case (see e.g. Laffont and Tirole (1987)).

6 Multi–Stage Contracting

As indicated in Section 2 we may apply our results also to multi–stage contracting problems. By the technique of dynamic programming such a problem can be stated as a sequence of static problems. Consider a situation

[12]Note that the previous example has been constructed in such a way that in the solution the agent's strategy assign's zero probability to the allocation $(s(m_3), w(m_3))$ in (12). One may easily check that the principal's payoff under the outcome $(q, s, w|M)$ in the previous example is $105/4$. It is thus Pareto dominated by the solution $(\tilde{q}, \tilde{s}, \tilde{w}|T)$.

[13]In our example the incentive constraint for type t_2 is binding with full commitment, whereas it is the incentive constraint for type t_1 that is binding with imperfect commitment. The latter observation, however, can be shown to depend on the specific parameters of the example.

in which the principal and the agent contract repeatedly over a sequence, $\tau = 1, ..., \bar{\tau}$, of periods. At each date τ the principal can commit himself to a decision $x_\tau \in X_\tau$; he cannot commit himself to decisions in future periods. We denote the decisions that have been implemented in the periods up to date τ by $\bar{x}_\tau \equiv (x_1, ..., x_\tau)$. The principal's and the agent's payoff in period τ depend on $\bar{x}_\tau$ and the agent's type t_i according to $v_{i\tau}(\bar{x}_\tau)$ and $u_{i\tau}(\bar{x}_\tau)$. We thus allow for the possibility that current decisions may affect the payoffs at later dates. The parties discount future payoffs with the discount factor $0 < \delta \leq 1$ so that, if the agent's type is t_i, their overall payoffs are given by

$$\sum\nolimits_{\tau=1}^{\bar{\tau}} \delta^{\tau-1} v_{i\tau}(\bar{x}_\tau), \quad \sum\nolimits_{\tau=1}^{\bar{\tau}} \delta^{\tau-1} u_{i\tau}(\bar{x}_\tau). \tag{28}$$

Let $\bar{U}_{i\tau}(\bar{x}_{\tau-1})$ denote the reservation utility of the t_i-type agent in period τ. The principal enters period τ with the belief $p_{\tau-1} \in P$. He chooses a contract $\Gamma_\tau = (M_\tau, x_\tau)$ so that $x_\tau \colon M_\tau \to X_\tau$ commits him to select $x_\tau(m)$ when the agent's reporting strategy $q_\tau \colon T \to Q_\tau$ selects the message $m \in M_\tau$.[14]

The principal faces a dynamic problem whose *state* at the beginning of date τ is described by $(\bar{x}_{\tau-1}, p_{\tau-1})$. Given a state, he uses the agent's message m to update his belief according to some function $p_\tau \colon M_\tau \to P$. This function together with the contract Γ_τ and the agent's reporting strategy q_τ determines the subsequent state $(\bar{x}_{\tau-1}, x_\tau, p_\tau)$ at date $\tau + 1$. We denote the *value functions* of the contracting problem by $v^*_{i\tau}(\bar{x}_{\tau-1}, p_{\tau-1})$ and $u^*_{i\tau}(\bar{x}_{\tau-1}, p_{\tau-1})$, with $v^*_{i\bar{\tau}+1} = u^*_{i\bar{\tau}+1} = 0$. Thus, if the agent's type is t_i, the functions $v^*_{i\tau}$ and $u^*_{i\tau}$ describe how the principal's and the agent's expected payoffs over the remainder of the contracting problem depend on the current state at the beginning of period τ. In what follows, let

$$\begin{aligned} V_{i\tau}(\bar{x}_\tau, p_\tau) &\equiv v_{i\tau}(\bar{x}_\tau) + \delta\, v^*_{i\tau+1}(\bar{x}_\tau, p_\tau), \\ U_{i\tau}(\bar{x}_\tau, p_\tau) &\equiv u_{i\tau}(\bar{x}_\tau) + \delta\, u^*_{i\tau+1}(\bar{x}_\tau, p_\tau). \end{aligned} \tag{29}$$

To state the principal's problem at date τ in state $(\bar{x}_{\tau-1}, p_{\tau-1})$, define for a given message set $M_\tau = M_\tau(\bar{x}_{\tau-1}, p_{\tau-1})$ the expected payoffs

$$\begin{aligned} V^*_{i\tau}(q_{i\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1} | M_\tau) &\equiv \int_{M_\tau} V_{i\tau}(\bar{x}_{\tau-1}, x_\tau(m), p_\tau(m))\, \mathrm{d}q_{i\tau}(m), \\ U^*_{i\tau}(q_{i\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1} | M_\tau) &\equiv \int_{M_\tau} U_{i\tau}(\bar{x}_{\tau-1}, x_\tau(m), p_\tau(m))\, \mathrm{d}q_{i\tau}(m). \end{aligned} \tag{30}$$

[14] This framework is directly applicable to short–term contracting in a repeated relationship, as e.g. in Laffont and Tirole (1987) or Schmidt (1993). By appropriately specifying the payoffs $v_{i\tau}(\cdot), u_{i\tau}(\cdot)$ and $\bar{U}_{i\tau}(\cdot)$, it can be adapted to study also the renegotiation of long–term contracts, as e.g. in Hart and Tirole (1988) and Laffont and Tirole (1990).

It follows that the principal's expected payoff at the beginning of period τ is

$$\sum_i p_{i\,\tau-1} V^*_{i\tau}(q_{i\,\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1}|M_\tau). \tag{31}$$

Given M_τ, the principal's objective in the state $(\bar{x}_{\tau-1}, p_{\tau-1})$ is to choose x_τ, q_τ and p_τ to maximize this payoff subject to three constraints: First, the agent selects his reporting strategy anticipating the effect of his message on the principal's belief. Therefore, for all $t_i \in T$ it has to be the case that

$$U^*_{i\tau}(q_{i\,\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1}|M_\tau) \geq U^*_{i\tau}(q'_i, p_\tau, x_\tau, \bar{x}_{\tau-1}|M_\tau) \quad \text{for all} \quad q'_i \in Q_\tau. \tag{32}$$

Second, the principal updates his belief according to Bayes' rule so that

$$\int_H p_{i\,\tau}(m) \mathrm{d}\bar{q}_\tau(m) = p_{i\,\tau-1} q_{i\,\tau}(H) \quad \text{for all} \quad H \in \mathcal{M}_\tau \text{ with } \bar{q}_\tau(H) > 0, \tag{33}$$

where $p_0 \equiv \gamma$. Finally, the agent's individual rationality constraint has to be satisfied for all types to which the principal assigns positive probability.[15] This condition can be stated as

$$\left[U^*_{i\tau}(q_{i\,\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1}|M_\tau) - \bar{U}_{i\,\tau}(\bar{x}_{\tau-1})\right] p_{i\tau-1} \geq 0 \quad \text{for all} \quad t_i \in T. \tag{34}$$

Let $(q^*_\tau,\, p^*_\tau,\, x^*_\tau)$ solve the principal's problem in the state $(\bar{x}_{\tau-1}, p_{\tau-1})$. Then it follows from the Principle of Optimality that

$$\begin{aligned} v^*_{i\,\tau}(\bar{x}_{\tau-1}, p_{\tau-1}) &= V^*_{i\tau}(q^*_{i\,\tau}, p^*_\tau, x^*_\tau, \bar{x}_{\tau-1}|M_\tau), \\ u^*_{i\,\tau}(\bar{x}_{\tau-1}, p_{\tau-1}) &= U^*_{i\,\tau}(q^*_{i\,\tau}, p^*_\tau, x^*_\tau, \bar{x}_{\tau-1}|M_\tau). \end{aligned} \tag{35}$$

Therefore, given a collection of (possibly state dependent) message sets $M_\tau(\cdot)$, $\tau = 1, ..., \bar{\tau}$, the value functions of the contracting problem can be derived recursively to solve the principal's multi–stage problem backwards from date $\tau = \bar{\tau}$ to $\tau = 1$.[16]

To apply Proposition 2 to the present context, we say that $(q_\tau, p_\tau, x_\tau|M_\tau)$ is *incentive feasible* if it satisfies (32) and (33). Further $(q_\tau, p_\tau, x_\tau|M_\tau)$ is *incentive efficient* if it is incentive feasible and there is no incentive feasible $(q'_\tau, p'_\tau, x'_\tau|M_\tau)$ such that

$$\begin{aligned} \sum_i p_{i\,\tau-1}[V^*_{i\tau}(q'_{i\,\tau}, p'_\tau, x'_\tau, \bar{x}_{\tau-1}|M_\tau) &- V^*_{i\tau}(q_{i\,\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1}|M_\tau)] > 0, \\ U^*_{i\,\tau}(q'_{i\,\tau}, p'_\tau, x'_\tau, \bar{x}_{\tau-1}|M_\tau) &= U^*_{i\,\tau}(q_{i\,\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1}|M_\tau), \end{aligned} \tag{36}$$

[15] As in Section 2, we assume for convenience that the principal wants to guarantee the participation of all types t_i with $p_{i\tau-1} > 0$.

[16] If $|\tau| = \infty$, one may derive the value functions by recursive methods as described in Stokey and Lucas (1989).

for all $t_i \in T$. Clearly, a solution to the principal's problem must be incentive efficient for any state $(\bar{x}_{\tau-1}, p_{\tau-1})$ in period τ that is reached with positive probability. Finally, $(q_\tau, p_\tau, x_\tau | M_\tau)$ and $(q'_\tau, p'_\tau, x'_\tau | M'_\tau)$ are *payoff–equivalent* if $\sum_i p_{i\,\tau-1} V^*_{i\tau}(q_{i\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1} | M_\tau) = \sum_i p_{i\,\tau-1} V^*_{i\tau}(q'_{i\tau}, p'_\tau, x'_\tau, \bar{x}_{\tau-1} | M'_\tau)$ and $U^*_{i\tau}(q_{i\tau}, p_\tau, x_\tau, \bar{x}_{\tau-1} | M_\tau) = U^*_{i\tau}(q'_{i\tau}, p'_\tau, x'_\tau, \bar{x}_{\tau-1} | M'_\tau)$ for all $t_i \in T$. With these modifications of our concepts in Section 2 we can prove the following Corollary to Proposition 2:

Corollary *Consider a state $(\bar{x}_{\tau-1}, p_{\tau-1})$ in period τ. If $(q_\tau, p_\tau, x_\tau | M_\tau)$ is incentive efficient, then there exists a direct mechanism $\Gamma^d_\tau = (T, x'_\tau)$ and an incentive feasible $(q'_\tau, p'_\tau, x'_\tau | T)$ such that $(q'_\tau, p'_\tau, x'_\tau | T)$ and $(q_\tau, p_\tau, x_\tau | M_\tau)$ are payoff–equivalent. Moreover, $q'_{i\tau}(t_i) > 0$ for all $t_i \in T$.*

The Corollary implies that in any state $(\bar{x}_{\tau-1}, p_{\tau-1})$ the principal cannot do better than apply a direct mechanism to solve the following maximization problem, which is analogous to (17) – (22):

$$
\begin{aligned}
\max_{q_\tau, p_\tau, x_\tau} \quad & \textstyle\sum_i \sum_j p_{i\,\tau-1} q_{i\tau}(t_j)\, V_{i\tau}\left(\bar{x}_{\tau-1}, x_\tau(t_j), p_\tau(t_j)\right) \qquad (37)\\
\text{s.t.} \quad & U_{i\tau}\left(\bar{x}_{\tau-1}, x_\tau(t_i), p_\tau(t_i)\right) \geq U_{i\tau}\left(\bar{x}_{\tau-1}, x_\tau(t_j), p_\tau(t_j)\right),\\
& \left[U_{i\tau}\left(\bar{x}_{\tau-1}, x_\tau(t_i), p_\tau(t_i)\right) - \bar{U}_{i\tau}(\bar{x}_{\tau-1})\right] p_{i\tau-1} \geq 0,\\
& \left[U_{i\tau}\left(\bar{x}_{\tau-1}, x_\tau(t_i), p_\tau(t_i)\right) - U_{i\tau}\left(\bar{x}_{\tau-1}, x_\tau(t_j), p_\tau(t_j)\right)\right] q_{i\tau}(t_j) = 0,\\
& p_{i\tau}(t_j) \textstyle\sum_k p_{k\,\tau-1}\, q_{k\tau}(t_j) = p_{i\,\tau-1}\, q_{i\tau}(t_j),
\end{aligned}
$$

for all $t_i, t_j \in T$. This program together with the recursive determination of the associated value functions in (35) reduces the problem of multi–stage mechanism design to a standard dynamic programming problem.

By the last condition of problem (37) the solution of the multi–stage contracting problem determines sequences of beliefs, $p_0, p_1, ..., p_{\bar{\tau}}$, that support the principal's choice of contract in each state as sequentially optimal behavior. As the Corollary indicates and our example shows, under imperfect commitment the optimal mechanism may induce the agent to reveal his information only partially. In multi–stage settings this means that the principal's beliefs are updated on the basis of a *gradual* revelation of information, as in Dewatripont (1989) and Hart and Tirole (1988).

7 Concluding Remarks

The contribution of this paper is an extension of the revelation principle to contracting problems with limited commitment. Our main result provides a useful tool for the study of such problems between a principal and a single

agent. It shows that the solution can be found in a subset of the set of incentive compatible allocations. By precisely stating the constraints that go beyond incentive compatibility, we derive a programming problem that can be used to study a wide variety of contracting situations with imperfect commitment.

As most of the literature on imperfect commitment, our analysis addresses contracting problems between a principal and a single agent. In Bester and Strausz (2000) we give an example which shows that a direct mechanism may no longer be optimal when the principal contracts with two agents. Therefore, it remains an open question to what extent one can characterize optimal mechanisms in settings with limited commitment and multiple agents. Green and Laffont's (1987) study of 'posterior implementable' rules makes a step in this direction by describing feasible agreements when no binding commitments are possible during the communication stage.

8 Appendix

Proof of Lemma 1: Let $K = \{K_1, ..., K_k, ...\}$ be a $\sigma-$partition of M. Then for any $\lambda = (\lambda_1, ..., \lambda_k, ...)$ such that

$$\lambda_k \geq 0 \text{ for all } K_k \in K, \quad \sum\nolimits_k \lambda_k \, q_i(K_k) = 1 \quad \text{for all } t_i \in T, \tag{38}$$

we can define a new strategy q' for the agent by setting

$$q_i'(H) \equiv \sum\nolimits_k \lambda_k \, q_i(H \cap K_k) \tag{39}$$

for all $t_i \in T$ and all $H \in \mathcal{M}$. Indeed, by (38), we have $q_i' \in Q$ for all $t_i \in T$. We now show that incentive feasibility of $(q, p, y, x|M)$ implies that also $(q', p, y, x|M)$ is incentive feasible. Notice that (3) implies

$$U_i(x(m), y(m)) = U_i^*(q, y, x|M) \quad q_i - \text{almost everywhere.} \tag{40}$$

Therefore,

$$U_i^*(q', y, x|M) = \sum\nolimits_k \lambda_k \int_{K_k} U_i(x(m), y(m)) \mathrm{d}q_i(m) = U_i^*(q, y, x|M). \tag{41}$$

Thus also q_i' maximizes the type t_i-agents expected payoff, which proves that $(q', p, y, x|M)$ satisfies condition (3). To show that $(q', p, y, x|M)$ satisfies condition (4) consider any $H \in \mathcal{M}$ and let $H_k \equiv H \cap K_k$. Then, by (4)

$$\int_{H_k} p_i(m) \mathrm{d}\bar{q} = \gamma_i q_i(H_k) \tag{42}$$

whenever $\bar{q}(H_k) > 0$. Therefore

$$\int_H p_i(m) \mathrm{d}\bar{q}\,' = \sum\nolimits_k \int_{H_k} p_i(m) \lambda_k \mathrm{d}\bar{q} = \sum\nolimits_k \gamma_i \lambda_k q_i(H_k) = \gamma_i q_i'(H) \tag{43}$$

so that consistency with Bayes' rule is satisfied. Finally, condition (2) is trivially satisfied for $(q', p, y, x|M)$ because it holds for $(q, p, y, x|M)$.

Given $(q', p, y, x|M)$, the principal gets the payoff

$$\sum\nolimits_k \lambda_k \sum\nolimits_i \gamma_i \int_{K_k} V_i(x(m), y(m)) \mathrm{d}q_i(m). \tag{44}$$

Incentive efficiency of $(q, p, y, x|M)$ implies that $\lambda' = (1, ..., 1, ...)$ maximizes this payoff subject to (38). Therefore, there exists a $\mu = (\mu_1, ..., \mu_i, ..., \mu_{|T|})$ such that λ' satisfies the first order condition

$$\sum\nolimits_i \gamma_i \int_{K_k} V_i(x(m), y(m)) \mathrm{d}q_i(m) = \sum\nolimits_i \mu_i q_i(K_k) \tag{45}$$

for all k. By (4), this condition is identical to

$$\sum_i \int_{K_k} V_i(x(m), y(m))\, p_i(m)\, \mathrm{d}\bar{q}(m) = \sum_i \frac{\mu_i}{\gamma_i} \int_{K_k} p_i(m)\, \mathrm{d}\bar{q}(m) \qquad (46)$$

for all $K_k \in K$. Since (46) must hold for any arbitrary $\sigma-$partition K of M, this proves the statement of the lemma. Q.E.D.

Proof of Proposition 1: Note that by conditions (3) and (4)

$$[U_i(x(m), y(m)) - U_i^*(q, y, x|M)]\; p_i(m) = 0 \quad \bar{q} - \text{almost everywhere.} \quad (47)$$

Indeed, suppose there is a $H \in \mathcal{M}$ such $\int_H p_i(m)\mathrm{d}\bar{q} > 0$ and $U_i(x(m), y(m)) \neq U_i^*(q, y, x|M)$ for all $m \in H$. Then condition (4) requires $q_i(H) > 0$. Yet, as condition (3) implies (40), we have $q_i(H) = 0$, a contradiction.

By Lemma 1 and (47) the support of $\bar{q}$ contains a set $\bar{M} \in \mathcal{M}$ with $\bar{q}(\bar{M}) = 1$ such that the following two properties are satisfied: First, there is $\mu \in \mathbb{R}^{|T|}$ such that

$$\sum\nolimits_i V_i(x(m), y(m)) p_i(m) = \sum\nolimits_i \mu_i\, p_i(m)/\gamma_i \quad \text{for all } m \in \bar{M}. \qquad (48)$$

Second,

$$[U_i(x(m), y(m)) - U_i^*(q, y, x|M)]\; p_i(m) = 0 \quad \text{for all } m \in \bar{M}. \qquad (49)$$

Moreover, since $\bar{q}(\bar{M}) = 1$ implies $q_i(\bar{M}) = 1$, it follows from (4) that

$$\int_{\bar{M}} p(m)\mathrm{d}\bar{q} = \gamma, \qquad (50)$$

where $p(m) = \left(p_1(m), ..., p_{|T|}(m)\right)$.

Define $\bar{P} = \{p(m)|m \in \bar{M}\}$ and let $\mathrm{co}(\bar{P})$ denote the convex hull of $\bar{P}$. By a theorem of Rubin and Wesler (1958), it follows from (50) that $\gamma \in \mathrm{co}(\bar{P})$. Since $\mathrm{co}(\bar{P})$ lies on the hyperplane $\left\{p \in \mathbb{R}^{|T|} | \sum_i p_i = 1\right\}$, it may be represented as a set in $\mathbb{R}^{|T|-1}$. Therefore, Caratheodory's theorem asserts that γ may be written as a convex combination of $|M'| \leq |T|$ linearly independent vectors $p(m_1), ..., p(m_{|M'|})$ in $\bar{P}$. Thus there exists $\alpha = (\alpha_1, ..., \alpha_h, ..., \alpha_{|M'|})$ such that $\alpha_h \geq 0$, $\sum_h \alpha_h = 1$ and

$$\sum\nolimits_h \alpha_h\, p(m_h) = \gamma. \qquad (51)$$

Consider the message set $M' = \{m_1, ..., m_h, ..., m_{|M'|}\}$ associated with the vectors $p(m_1), ..., p(m_h), ..., p(m_{|M'|})$ and define for all $H \in \mathcal{M}$ the agent's reporting strategy q' by

$$q_i'(H) = \sum\nolimits_{m_h \in H} \frac{\alpha_h}{\gamma_i}\, p_i(m_h) \qquad (52)$$

By (51), $q_i' \in Q$. The vectors $q'(m_h)$, $h = 1, ..., |M'|$, are linearly independent because the vectors $p(m_h)$, $h = 1, ..., |M'|$, are linearly independent.

We now show that $(q', p, y, x|M)$ is incentive feasible. First, condition (2) is trivially satisfied because $(q', p, y, x|M)$ and $(q, p, y, x|M)$ differ only in the agent's strategy. Second we have $q_i'(M') = 1$ and $q_i'(m_h) > 0$ only if $p_i(m_h) > 0$. Since $M' \subset \bar{M}$, this together with (49) implies that

$$\sum\nolimits_h U_i(x(m_h), y(m_h)) q_i'(m_h) = U_i^*(q, y, x|M) \geq \int_M U_i(x(m), y(m)) \mathrm{d}q_i''(m) \quad \text{for all} \quad q_i'' \in Q. \tag{53}$$

This proves that $(q', p, y, x|M)$ satisfies condition (3). Finally, $(q', p, y, x|M)$ satisfies condition (4) because $\bar{q}\,'(M') = 1$ and

$$p_i(m_h) \sum\nolimits_j \gamma_j q_j'(m_h) = q_i'(m_h) \frac{\gamma_i}{\alpha_h} \sum\nolimits_j \alpha_h p_j(m_h) = \gamma_i\, q_i'(m_h) \tag{54}$$

for all $m_h \in M'$.

It remains to show that $(q', p, y, x|M)$ and $(q, p, y, x|M)$ are payoff–equivalent. By (53) we have $U_i^*(q, y, x|M) = U_i^*(q', y, x|M)$ for all $t_i \in T$. Suppose that $V^*(q, y, x|M) \neq V^*(q', y, x|M)$. Because $(q, p, y, x|M)$ is incentive efficient, this implies $V^*(q, y, x|M) > V^*(q', y, x|M)$. Therefore, using (4), we obtain by Lemma 1 and (48) that

$$\begin{aligned} &\sum\nolimits_i \int_M V_i(x(m), y(m))\, p_i(m)\, \mathrm{d}\bar{q}(m) = \sum\nolimits_i \int_M \frac{\mu_i}{\gamma_i} p_i(m)\, \mathrm{d}\bar{q}(m) > \\ &\sum\nolimits_i \int_M \frac{\mu_i}{\gamma_i} p_i(m)\, \mathrm{d}\bar{q}\,'(m) = \sum\nolimits_i \int_M V_i(x(m), y(m))\, p_i(m)\, \mathrm{d}\bar{q}\,'(m), \end{aligned} \tag{55}$$

because $\bar{q}\,'(\bar{M}) = \bar{q}\,'(M') = 1$. Since by condition (4) $\int_M p_i(m)\, \mathrm{d}\bar{q}(m) = \gamma_i$ $= \int_M p_i(m)\, \mathrm{d}\bar{q}\,'(m)$, the inequality in (55) cannot hold. This contradiction proves that $(q', p, y, x|M)$ and $(q, p, y, x|M)$ are payoff–equivalent. Q.E.D.

Proof of Proposition 2: Because we can simply delete from M any $H \in \mathcal{M}$ such that $\bar{q}\,'(H) = 0$, Proposition 1 guarantees that there exists a mechanism (M', x') with $|M'| \leq |T|$ and an incentive feasible $(q', p', y', x'|M')$ which is payoff–equivalent to $(q, p, y, x|M)$. Let $\Omega_{M'} = [q'(m_h)]_{m_h \in M'}$ denote the $|T| \times |M'|$ matrix with column vectors $q'(m_h), h = 1, ..., |M'|$. Note that by Proposition 1 the column vectors of $\Omega_{M'}$ are linearly independent.

Define the correspondence $D\colon M' \Rightarrow T$ by $D(m_h) = \{t_i | q_i'(m_h) > 0\}$. It follows that $D(H) = \cup_H \{t_i | q_i'(m_h) > 0, m_h \in H\}$. We claim that $|D(H)| \geq |H|$ for all $H \subseteq M'$. Indeed, fix $H \subseteq M'$ and consider the $|T| \times |H|$ matrix

$\Omega_H = [q'(m_h)]_{m_h \in H}$. Since the matrix $\Omega_{M'}$ consists of linearly independent column vectors, this also holds for Ω_H. Consequently, $\text{Rank}(\Omega_H) = |H|$. Note further that the matrix Ω_H has only $|D(H)|$ non–null row vectors. This implies that $\text{Rank}(\Omega_H) \leq |D(H)|$. Hence, it follows that $|D(H)| \geq |H|$. Lemma 2 therefore asserts the existence of a mapping $d\colon M' \to T$ with $d(m_h) \in D(m_h)$ and the property that $d(m_h) = d(m_k)$ implies $m_h = m_k$.

We now use $d(\cdot)$ to construct a mapping $c\colon T \to M'$ in the following way: Since the mapping $d(\cdot)$ is invertible we can set $c(d(m_h)) = m_h$ for each $m_h \in M'$. As $d(m_h) \in D(m_h)$ we have $q_i'(c(t_i)) > 0$ for all $t_i \in T^o \equiv \{d(m_h)|m_h \in M'\}$. To each $t_i \notin T^o$ we can assign an arbitrary $c(t_i) \in M'$ such that $q_i'(c(t_i)) > 0$. Such a $c(t_i)$ exists because $\sum_h q_i'(m_h) = 1$. Thus the mapping $c(\cdot)$ satisfies $q_i(c(t_i)) > 0$ for all $t_i \in T$. Moreover, as $\cup_{T^o} c(t_i) = M'$ we have that

$$S(m_h) \equiv \{t_i | m_h = c(t_i)\} \neq \emptyset \tag{56}$$

for all $m_h \in M'$.

Now we replace the mechanism (M', x') and $(q', p', y', x'|M')$ by a direct mechanism $(T, \hat{x})$ and a $(\hat{q}, \hat{p}, \hat{y}, \hat{x}|T)$ that is defined in the following way:

$$\hat{q}_i(t_j) = \frac{q_i'(c(t_j))}{|S(c(t_j))|},\ \hat{p}(t_j) = p'(c(t_j)),\ \hat{y}(t_j) = y'(c(t_j)),\ \hat{x}(t_j) = x'(c(t_j)). \tag{57}$$

Note that $\hat{q}_i(t_i) > 0$ for all $t_i \in T$. Thus, to complete the proof, it is sufficient to show that $(\hat{q}, \hat{p}, \hat{y}, \hat{x}|T)$ is incentive feasible and payoff–equivalent to $(q', p', y', x'|M')$.

By (57), $\hat{q}_i(t_j) = q_i'(m_h)/|S(m_h)|$ for all $t_j \in S(m_h)$. Therefore,

$$\sum\nolimits_j \hat{q}_i(t_j) = \sum\nolimits_h \sum\nolimits_{t_j \in S(m_h)} q_i'(m_h)/|S(m_h)| = \sum\nolimits_h q_i'(m_h) = 1, \tag{58}$$

so that $\hat{q}_i$ defines a probability distribution on T. Since $(\hat{x}(t_j), \hat{y}(t_j)) = (x'(c(t_j)), y'(c(t_j)))$, any allocation that the agent induces by some message $t_j \in T$ under the mechanism $(T, \hat{x})$ he can also induce by the message $c(t_j) \in M$ under the mechanism (M', x'). Conversely, as for each $m_h \in M$ there is a $t_i \in T$ such that $m_h = c(t_i)$, anything that he can induce under (M', x') he can also induce under $(T, \hat{x})$. Therefore $U_i^*(\hat{q}, \hat{y}, \hat{x}|T) = U_i^*(q', y', x'|M')$ for all $t_i \in T$. Moreover, $\hat{q}_i(t_j) > 0$ if and only if $q_i'(c(t_j)) > 0$. Thus $\hat{q}$ satisfies condition (3).

The principal's belief $\hat{p}$ is consistent with condition (4) because

$$\hat{p}_i(t_j) = \frac{\gamma_i \hat{q}_i(t_j)}{\sum_k \gamma_k \hat{q}_k(t_j)} = \frac{\gamma_i q_i'(c(t_j))}{\sum_k \gamma_k q_k'(c(t_j))} = p_i'\left(c(t_j)\right). \tag{59}$$

Since $\hat{p}(t_j) = p'(c(t_j))$ and $\hat{x}(t_j) = x'(c(t_j))$, the principal's choice $\hat{y}(t_j) = y'(c(t_j))$ satisfies condition (2). Finally, under $(q', p', y', x'|M')$ the t_i-type agent induces the allocation $(x'(m_h), y'(m_h))$ with probability $q'_i(m_h)$. Under $(\hat{q}, \hat{p}, \hat{y}, \hat{x}|T)$ he induces the same allocation with the same probability, as $\sum_{t_j \in S(m_h)} \hat{q}_i(t_j) = q'_i(m_h)$. Therefore, $V^*(\hat{q}, \hat{y}, \hat{x}|T) = V^*(q', y', x'|M')$. Q.E.D.

Proof of the Corollary: To apply the proofs of Lemma 1 and of Propositions 1 and 2, replace (q, p, y, x) by (q_τ, p_τ, x_τ) and the functions $U_i(\cdot)$, $V_i(\cdot)$, $U_i^*(\cdot)$, $V^*(\cdot)$ by $U_{i\,\tau}(\cdot), V_{i\,\tau}(\cdot), U^*_{i\tau}(\cdot), \sum_i p_{i\,\tau-1} V^*_{i\tau}(\cdot)$. Further replace γ by $p_{\tau-1}$ and note that instead of (2) – (4) the incentive feasibility conditions (32) and (33) apply.

Consider a state $(\bar{x}_{\tau-1}, p_{\tau-1})$. If $p_{i\tau-1} > 0$ for all $t_i \in T$, the Corollary follows directly from the arguments leading to Proposition 2. If $p_{i\tau-1} = 0$ for some $t_i \in T$, we proceed in two steps: First, we apply Proposition 2 to the reduced type space $T' \equiv \{t_i \in T |\, p_{i\tau-1} > 0\}$ to show that there is a direct mechanism (T', x'_τ) and an incentive feasible $(q'_\tau, p'_\tau, x'_\tau |\, T')$ that is payoff–equivalent and satisfies $q'_i(t_i) > 0$ for all $t_i \in T'$.

Second, to extend $(q'_\tau, p'_\tau, x'_\tau)$ to the original type space T, select for every $t_i \in T \backslash T'$ some message $m_i \in M_\tau$ in the support of q_i that maximizes $U_{i\,\tau}(\bar{x}_{\tau-1}, x_\tau(\cdot), p_\tau(\cdot))$. Define $x'_\tau(t_i) = x_\tau(m_i)$ and $p'_\tau(t_i) = p_\tau(m_i)$; further define $q'_i(t_i) = 1$ and $q'_j(t_i) = 0$ for all $t_j \neq t_i$. Thus $q'_i(t_i) > 0$ for all $t_i \in T$. By construction, $(q'_\tau, p'_\tau, x'_\tau |\, T)$ satisfies the incentive feasibility conditions (32) and (33). In addition, $(q'_\tau, p'_\tau, x'_\tau |\, T)$ is payoff–equivalent to $(q_\tau, p_\tau, x_\tau | M_\tau)$. Q.E.D.

9 References

Bester, H. and R. Strausz (2000): "Imperfect Commitment and the Revelation Principle: The Multi–Agent Case," *Economics Letters*, forthcoming.

Crawford, V. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica 50*, 1431–1451.

Dasgupta, P., P. Hammond and E. Maskin (1979): "The Implementation of Social Choice Rules," *Review of Economic Studies 46*, 185–216.

Dewatripont, M. (1989): "Renegotiation and Information Revelation over Time: The Case of Optimal Labor Contracts," *Quarterly Journal of Economics 104*, 589–619.

Epstein, L. G. and M. Peters (1996): "A Revelation Principle for Competing Mechanisms," Working Paper # UT–ECIPA-PETERS-02, Dept. of Economics, University of Toronto.

Freixas, X., R. Guesnerie and J. Tirole (1985): "Planning under Incomplete Information and the Ratchet Effect," *Review of Economic Studies 52*, 173–191.

Gibbard, A. (1973): "Manipulation for Voting Schemes," *Econometrica 41*, 587–601.

Green, J. and J.–J. Laffont (1977): "Characterization of Satisfactory Mechanisms for the Revelation of Preferences," *Econometrica 45*, 427–438.

Green, J. and J.–J. Laffont (1987): "Posterior Implementability in a Two–Person Decision Problem," *Econometrica 55*, 69–94.

Hart, O. and J. Tirole (1988): "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies 55*, 509–540.

Hall, Ph. (1935): "On Representatives of Subsets," *Journal of the London Mathematical Society 10*, 26–30.

Jacobs, K. (1969): *Selecta Mathematica I*, HTB 49, Springer–Verlag: Berlin, Heidelberg, New York.

Khalil, F. (1997): "Auditing without Commitment," *RAND Journal of Economics 28*, 629–640.

Laffont, J.–J. and J. Tirole (1987): "Comparative Statics of the Optimal Dynamic Incentive Contract," *European Economic Review 31*, 901–926.

Laffont, J.–J. and J. Tirole (1990): "Adverse Selection and Renegotiation in Procurement," *Review of Economic Studies 57*, 579–625.

Laffont, J.–J. and J. Tirole (1993): *A Theory of Incentives in Procurement and Regulation*, MIT Press: Cambridge – Massachusetts, London.

Maak, W. (1936): "Eine neue Definition der fastperiodischen Funktionen," *Abhandlungen aus dem Mathematischen Seminar Hamburg 11*, 240–244.

Martimort, D. and L. Stole (1999): "The Revelation and Taxation Principles in Common Agency Games," University of Chicago, *mimeo.*

Maskin, E. and J. Tirole (1999): "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies 66*, 83–114.

Miyazaki, H. (1977): "The Rat Race and Internal Labor Markets," *Bell Journal of Economics 8*, 394–418.

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

Rockafellar, R. T. (1970): *Convex Analysis,* Princeton University Press, Princeton NJ.

Rubin, H. and O. Wesler (1958): "A Note on Convexity in Euclidean n-Space", *Proceedings of the American Mathematical Society 9*, 522–523.

Schmidt, K. M. (1993): "Commitment through Incomplete Information in a Simple Repeated Bargaining Game," *Journal of Economic Theory 60*, 114–139.

Stokey, N. L. and R. E. Lucas (1989): *Recursive Methods in Economic Dynamics,* Harvard University Press, Cambridge MA.

Weyl, H. (1949): "Almost Periodic Invariant Vector Sets in a Metric Vector Space," *American Journal of Mathematics 71*, 178–205.