BONN ECON DISCUSSION PAPERS

Discussion Paper 21/2002

Mechanism Design by an Informed Principal - Pure-Strategy Equilibria for a Common Value Model

by

Rolf Tisljar

September 2002

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is
sponsored by the

Deutsche Post World Net

MAIL EXPRESS LOGISTICS FINANCE

Mechanism Design by an Informed Principal – Pure-Strategy Equilibria for a Common Value Model

Rolf Tisljar*
Bonn Graduate School of Economics
University of Bonn

September, 2002

Abstract

We present a common value mechanism design model for an informed principal where only the principal has private information, but her one-dimensional private information is allowed to be distributed according to any probability measure. For this model we characterize the set of pure-strategy perfect Bayesian equilibria. Furthermore, we present several equilibrium refinements based on the concept of equilibrium domination to take account of beliefs off the equilibrium path. Finally, we demonstrate that the extension of the *strong solution* of Myerson (Econometrica, 1983) to our model is supported as an equilibrium satisfying all refinement criteria presented (in case a strong solution exists).

JEL-Classification: C72, D82

Keywords:

equilibrium refinement, infinite signaling game, informed principal, mechanism design, perfect Bayesian equilibrium, principle of inscrutability, revelation principle, strong solution

*I would like to thank Georg Nöldeke, Clemens Puppe, Urs Schweizer, Thomas Wiseman, and participants at the Bonn Graduate School of Economics workshop for very helpful comments on previous versions of this paper. Financial support from the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

Address: Bonn Graduate School of Economics, University of Bonn,
Adenauerallee 24–42, 53113 Bonn, Germany
eMail: tisljar@wiwi.uni-bonn.de

1 Introduction

The theory of standard mechanism design for an uninformed principal has been well developed already over 20 years ago, and it has been applied to various economic problems like, for example, the problem of optimal auction design ([15] Myerson (1981)) or the problem of optimal regulation of a monopolist ([1] Baron and Myerson (1982)). However, the assumption that the principal does not have any relevant private information in these models is often rather artificial and seems to be driven by the desire to evade the difficulties associated with informed principal mechanism design. An auctioneer, for example, typically possesses some private information about the item to be sold, and a regulator might have some private information about, for example, the extent to which she cares about consumers' surplus and firms's profits. Similarly, an employer typically has private information about the job which is on offer when designing job contracts, and a big insurance customer has private information about her accident proneness when designing insurance contracts for an insurance company[1].

The number of papers in the literature dealing with such informed principal models is still fairly restricted, and, in fact, most papers analyzing models with an informed principal are not really considering general mechanism design models. Instead, the set of feasible contracts is often restricted exogenously such that the execution of contracts does not depend on some action taken by the principal and, thus, the mechanism design model actually reduces to a signaling model (see, for example, [3] Beaudry (1994) or [17] Ottaviani and Prat (2001)). This negligence of general informed principal mechanism design models may partly be due to the fact that so far solution concepts have been presented only for mechanism design models where the type of the principal is drawn from a finite set; see [16] Myerson (1983) for an analysis of a finite-type model with adverse selection and moral hazard, [12] Maskin and Tirole (1990) for a *private value* model with adverse selection only, and [13] Maskin and Tirole (1992) for a *common value* model with adverse selection.

The aim of this paper is to extend the analysis of [13] Maskin and Tirole (1992) by analyzing a common value mechanism design model for an informed principal where types are drawn in accordance with some general

[1]For more examples of situations where a privately informed principal designs mechanisms see [13] Maskin and Tirole (1992).

probability measure (so in particular we allow for types being distributed continuously). Maskin and Tirole have shown for their finite-type model (among other results) that each perfect Bayesian equilibrium (PBE) is given by an incentive-compatible allocation that is ex-ante individual rational for the agent and that weakly Pareto dominates the *Rothschild-Stiglitz-Wilson* allocation associated with the status quo. We replicate this result for the model with our more general information structure[2]: The main theorem of this paper demonstrates that a revelation and inscrutability theorem holds in our mechanism design model if the players' strategies are restricted in a suitable manner. The revelation part of the theorem then yields the incentive-compatibility as a necessary requirement, and the individual rationality result is easily derived from the inscrutability part[3]. Finally, it is easy to show that any equilibrium has to Pareto dominate the Rothschild-Stiglitz-Wilson allocation weakly because this allocation is almost by definition a feasible option for the principal.

Given that we can focus on direct incentive-compatible mechanisms, the great advantage of continuous-type over finite-type models in mechanism design is the fact that in continuous-type direct revelation mechanisms we can integrate over the incentive-compatibility restrictions of informed parties, which may help to characterize the set of such mechanisms, and hence simplify the description of equilibria (whereas in discrete-type models incentive-compatibility restrictions tend to make the characterization of solutions rather more than less difficult, and this complication may increase with the cardinality of the type space). [21] Tan (1996), for example, simplified the computation of equilibria for his model of optimal procurement design by an informed buyer by integrating over incentive-compatibility restrictions. Similarly, [19] Stoughton and Talmor (1994), [20] Stoughton and Talmor (1999), and [23] Yilankaya (1999) consider mechanism design mod-

[2]In addition to allowing for a more general information structure, our model is also more general than the model of [13] Maskin and Tirole (1992) with respect to the set of feasible allocations (D in our model) as well as the number of agents in the model.

[3][16] Myerson (1983) has pointed out that the Revelation Principle, which is well-established for mechanism design models with an uninformed principal, extends to his informed principal model. Though we do not prove the Revelation Principle separately, our revelation and inscrutability theorem like the Revelation Principle establishes that we can restrict ourselves to direct revelation mechanisms. Moreover, Myerson argues that in his finite type model a *Principle of Inscrutability* holds. The revelation and inscrutability theorem extends this principle to our mechanism design model.

els for an informed principal where the type of the principal is assumed to be continuously distributed, and they use differential representations of incentive-compatibility constraints. Note that all these papers employ solution concepts based on [16] Myerson (1983) or [13] Maskin and Tirole (1992), even though strictly speaking these solution concepts for finite-type models do not apply to the papers mentioned above. The analysis of our mechanism design model can be viewed as a first step in fixing this irregularity and in smoothing the way for the analysis of further informed principal mechanism design models (with continuously distributed private information).

The paper is organized as follows: In Section 2 we present the mechanism proposal, evaluation and execution game to be analyzed. Section 3 characterizes the set of allocation rules (or social choice functions) for this contract proposal game which can be implemented through pure-strategy perfect Bayesian equilibria (satisfying some regularity condition). The set of such equilibrium allocation rules may be relatively large, as PBE may involve unreasonable beliefs off the equilibrium path. To eliminate some of the unreasonable PBE, we present in Section 4 various equilibrium refinements based on the concept of equilibrium domination. Finally, in Section 5, we demonstrate that the extension to our model of the *strong solution* of [16] Myerson (1983) is supported as a PBE passing all the refinement criteria discussed – in case a strong solution exists. We conclude with a short summary of the paper. The proofs of the paper are presented in the appendix.

2 The Contract Proposal Game, Strategies, and Perfect Bayesian Equilibria

In this section we shall describe the *contract proposal game* to be analyzed, strategies and beliefs for this game, and the conditions for a pure-strategy perfect Bayesian equilibrium.

2.1 The Contract Proposal Game

In our model there are $n+1$ players – one principal, indexed by $i = 0$, and n agents, indexed by $i = 1, \ldots, n$. We assume that the agents do not possess any private information, but the principal has some private information represented by some $t \in T$, where T is some Borel set on the real line (e.g. T

some interval or some countable set in $\mathbb{R}$). Let $\mathcal{B} := \{B \subseteq T \mid B \text{ a Borel set}\}$, and let t be distributed according to some probability measure $P : \mathcal{B} \to [0,1]$, i.e. for all $B \in \mathcal{B}$ the probability that $t \in B$ is given by $P(B)$. So in particular we allow for types being distributed discretely as well as for types being distributed continuously.

The principal is to select a game form (or mechanism) m from some set of feasible game forms M which, if played by the $n+1$ players, determines an element d from a pre-specified set D of possible allocations (each $d \in D$ may specify some transfer payment and some output level, for example). Player i's von Neumann-Morgenstern utility derived from some $d \in D$ is given by $u_i(d,t) \in \mathbb{R}$ ($\forall\, t \in T$). So i's utility depends on the allocation $d \in D$ and on the principal's type (Common Values). We assume that for all $d \in D$, $i = 1, \ldots, n$, the function $t \mapsto u_i(d,t)$ is measurable[4]. Furthermore, let there be some $U^{max} \in \mathbb{R}$ such that $|u_i(d,t)| \leq U^{max}\ \forall\, i \in \{1, \ldots, n\},\ d \in D,\ t \in T$. The status quo is given by some $d_0 \in D$, and $U_i^0 := \int_T u_i(d_0,t)\mathrm{d}P(t)$ is the ex-ante expected utility agent i derives from d_0.

A feasible game form $m \in M$ is a deterministic function from some set $S_m := S_m^0 \times S_m^1 \times \ldots \times S_m^n$ to the set D, where the set S_m^i has to be chosen from some pre-specified set $\mathcal{S}^i \neq \emptyset$ $(i = 0, \ldots, n)$[5]. So $M \subseteq \{m : S_m^0 \times S_m^1 \times \ldots \times S_m^n \to D \mid S_m^i \in \mathcal{S}^i\ \forall\, i = 0, \ldots, n\}$. Hence a game form m represents a simultaneous-move game for the $n+1$ players, where player i has to choose an action $s^i \in S_m^i$ $(i = 0, \ldots, n)$, and payoffs are given by $u_i(m(s),t)$ for all $s = (s^0, s^1, \ldots, s^n) \in S_m$ $(i = 0, \ldots, n)$. By m we shall denote both the game form and the function $m : S_m^0 \times S_m^1 \times \ldots \times S_m^n \to D$ itself. If, for example, $\mathcal{S}^0 = \{T\}$ and $\mathcal{S}^i = \{\{1\}\}$ $(i = 1, \ldots, n)$, then M represents (some subset of) all direct mechanisms $m : T \times \{1\} \times \ldots \times \{1\} \to D$ (so basically all functions mapping claims about the principal's type to allocations in D). Finally, for later use we let M be endowed with some σ-algebra (also called σ-field) $\mathcal{M}$ (on M).

Now consider the following mechanism proposal, evaluation, and execution game, the *contract proposal game*[6]:

---

[4]So we assume that the function $f : (T, \mathcal{B}) \to \mathbb{R}$, $t \mapsto u_i(d,t)$, is measurable, where $\mathbb{R}$ is endowed with the Borel σ-algebra $\{B \subseteq \mathbb{R} \mid B \text{ a Borel set}\}$.

[5]Instead of requiring that m is a deterministic function, we could allow for functions m mapping into probability distributions over the set D and then take expected values.

[6]The structure of the game is the same as the contract proposal game analyzed by [13] Maskin and Tirole (1992).

1. Nature draws the principal's type $t \in T$; t is revealed to the principal, but all agents only know the probability measure according to which t is distributed.

2. The principal chooses a mechanism $m \in M$.

3. The n agents update their prior believe about the principal's type based on any information revealed by the selection of m.

4. The n agents simultaneously decide to accept or to reject the mechanism chosen by the principal, i.e. agent i either accepts ($a_i = 1$) or rejects ($a_i = 0$) the game form to be played.

5. If all agents have accepted the mechanism, the game proposed by the principal is played, i.e. the principal and the agents simultaneously choose an action $s_i \in S_m^i$ ($i = 0, \ldots, n$).

6. The payoffs to all players are given by

$$U_i(m, a, s, t) = \begin{cases} u_i(m(s), t) & : \quad a_i = 1 \; \forall \; i = 1, \ldots, n \\ u_i(d_0, t) & : \quad a_i = 0 \text{ for some } i > 0 \end{cases}$$

$\forall \; m \in M, \; a \in A := \{0, 1\}^n, \; s \in S_m, \; t \in T.$

So after the principal has observed her type, she announces some game (mechanism) to be played to determine an allocation $d \in D$ (the mechanism proposal stage). The agents, however, can reject the principal's mechanism (in the mechanism evaluation stage of the contract proposal game) – they have the possibility to choose action $a_i = 0$ and thereby secure themselves the status quo utility. It has to be pointed out that *all* agents have to accept the game to be played; a single veto will leave the status quo unchanged[7]. If a mechanism has been accepted, the game is played to determine the allocation to be implemented (the mechanism execution stage). We assume

[7] So we exclude the possibility that the principal could announce another game to be played, once a mechanism has been rejected. But such a situation would be interesting only if the agents had private information, too.
Moreover, it has to be pointed out here that allowing for more than one agent causes some conceptual problems: Rejecting a game form may be optimal just because other agents reject it as well, even though all agents would actually prefer playing the game as opposed to settling with the status quo. For some of our results we have to assume that this does not happen (or, alternatively, that $n = 1$).

that all agents commit themselves to accepting this allocation, even if it turns out ex-post that some agent would have preferred the status quo. The game to be played in the execution stage of the contract proposal game is a game of incomplete information, where we assume that the probability assessment for the principal's type is given by the updated belief that results from step three of the contract proposal game.

2.2 Pure Strategies and Updated Beliefs

Next we shall describe feasible pure strategies and admissible updated beliefs for the contract proposal game.

A pure strategy for the principal consists of a function $\sigma_0^1 : T \to M$ mapping types into mechanisms, and a function $\sigma_0^2 : T \times M \times A \to \bigcup_{S^0 \in \mathcal{S}^0} S^0$ (with $\sigma_0^2(t, m, 1^n) \in S_m^0 \;\forall\; m \in M,\; t \in T$, where $1^n = (1, \ldots, 1)$) specifying an action for the principal for the mechanism execution stage of the contract proposal game for all possible types. We require σ_0^2 to be such that each agent can calculate his expected utility from equilibrium play in the mechanism execution stage for all mechanisms $m \in M$[8].

For each agent $i = 1, \ldots, n$, a pure strategy is a function $\sigma_i^1 : M \to \{0, 1\}$ specifying if agent i accepts ($\sigma_i^1(m) = 1$) or rejects ($\sigma_i^1(m) = 0$) mechanism $m \in M$, and a function $\sigma_i^2 : M \times A \to \bigcup_{S^i \in \mathcal{S}^i} S^i$ (with $\sigma_i^2(m, 1^n) \in S_m^i \;\forall\, m \in M$).

Finally, agent i's updated belief concerning the principal's type after observing the announcement of a mechanism is given by a function $\beta_i : M \to \mathcal{P}(T)$ ($i = 1, \ldots, n$) mapping mechanisms into probability measures over the set T.

2.3 Pure-Strategy Perfect Bayesian Equilibria

A perfect Bayesian equilibrium (PBE) for the contract proposal game consists of a set of strategies and beliefs which are sequentially rational for all players and Bayesian consistent. So for a PBE we require that the principal's strategy for the mechanism proposal stage is optimal given the strategies of the agents, the agents' decisions concerning acceptance or rejection in the mechanism evaluation stage are optimal given their beliefs and the strategies of the other players, the beliefs of the agents formed after a mechanism

[8]Due to our restriction on the functions u_i it is always possible to find such strategies σ_0^2, e.g. all constant functions $\sigma_0^2(\cdot, m, 1^n)$ satisfy this restriction.

is proposed are Bayesian consistent, and the strategies of all players for the mechanism execution stage form a Bayesian Nash equilibrium, where the beliefs of the agents are the same as those used for the mechanism evaluation stage. Thus, a pure-strategy perfect Bayesian equilibrium is a set of strategies and beliefs $(\sigma^1, \sigma^2, \beta)$ which satisfy the following conditions:

1. **Measurability**
   σ^2 is such that the function

   $$t \mapsto u_i(m[\sigma_0^2(t, m, 1^n), \sigma_1^2(m, 1^n), \ldots, \sigma_n^2(m, 1^n)], t)$$

   is measurable for all $i \in \{1, \ldots, n\},\ m \in M$. Due to our boundedness assumption on the agents' utilities, this measurability assumption guarantees that $t \mapsto u_i(m[\sigma_0^2(t, m, 1^n), \sigma_1^2(m, 1^n), \ldots, \sigma_n^2(m, 1^n)], t)$ is integrable with respect to any updated belief $\beta_i(m)$ for any mechanism $m \in M$, and hence that all agents can calculate their expected utilities from equilibrium play in the mechanism execution stage for all $m \in M$.

2. **Sequential Rationality for the Principal**
   Sequential rationality for the principal reduces to the following two conditions:

   (a) $u_0(\sigma_0^1(t)[\sigma_0^2(t, \sigma_0^1(t), 1^n), \sigma_1^2(\sigma_0^1(t), 1^n), \ldots, \sigma_n^2(\sigma_0^1(t), 1^n)], t) \geq$
   $u_0(m[\sigma_0^2(t, m, 1^n), \sigma_1^2(m, 1^n), \ldots, \sigma_n^2(m, 1^n)], t)$
   $\forall\ t \in T,\ m \in M$ with $\sigma_i^1(m) = 1\ \forall\ i = 1, \ldots, n$;
   the principal's equilibrium utility has to be at least as large as the utility she could achieve by proposing some other mechanism in M which would be accepted by all agents. Moreover, the principal's utility has to be at least $u_0(d_0, t)$, since she can always secure herself her status quo utility (at least if the status quo is a feasible option to be announced by the principal or if there exists a mechanism which is guaranteed to be rejected by at least one agent). But this condition is implied by the requirement above, unless the game form which implements the status quo irrespective of the players' actions is not accepted by all agents (in which case we have to require explicitly that the principal does not end up with less than her status quo utility).

(b) $u_0(m[\sigma_0^2(t,m,1^n),\sigma_1^2(m,1^n),\ldots,\sigma_n^2(m,1^n)],t) \geq$
$u_0(m[s_0,\sigma_1^2(m,1^n),\ldots,\sigma_n^2(m,1^n)],t)$
$\forall\ t \in T,\ m \in M,\ s_0 \in S_m^0$;
this condition guarantees the optimality of the principal's action in the mechanism execution stage.

Note that – like in finite games – to check for optimality it is sufficient to compare the principal's equilibrium strategy to all other available *pure* strategies. This comparison has to be carried out at all information sets of the game tree where the player might be called on to take an action[9].

3. **Sequential Rationality for Agent** i
Sequential rationality for agent $i \in \{1,\ldots,n\}$ puts the following three restrictions on $(\sigma^1,\sigma^2,\beta)$:

(a) $\int_T u_i(m[\sigma_0^2(t,m,1^n),\sigma_1^2(m,1^n),\ldots,\sigma_n^2(m,1^n)],t)\mathrm{d}\beta_i(m)(t) \geq$
$\int_T u_i(d_0,t)\mathrm{d}\beta_i(m)(t)$
$\forall\ m \in M$ satisfying $\sigma_j^1(m) = 1$ for all $j = 1,\ldots,n$;
this condition ensures that accepting mechanism m is indeed optimal for agent i (given that all other agents accept m).

(b) $\int_T u_i(m[\sigma_0^2(t,m,1^n),\sigma_1^2(m,1^n),\ldots,\sigma_n^2(m,1^n)],t)\mathrm{d}\beta_i(m)(t) \leq$
$\int_T u_i(d_0,t)\mathrm{d}\beta_i(m)(t)$
for all $m \in M$ satisfying $\sigma_j^1(m) = 1\ \forall\ j \neq i,\ \sigma_i^1(m) = 0$;
this condition ensures that rejecting mechanism m is optimal for agent i (given that all other agents accept m)[10].

(c) $\int_T u_i(m[\sigma_0^2(t,m,1^n),\sigma_1^2(m,1^n),\ldots,\sigma_n^2(m,1^n)],t)\mathrm{d}\beta_i(m)(t) \geq$
$\int_T u_i(m[\sigma_0^2(t,m,1^n),\sigma_1^2(m,1^n),\ldots,\sigma_{i-1}^2(m,1^n),s_i,$
$\sigma_{i+1}^2(m,1^n),\ldots,\sigma_n^2(m,1^n)],t)\mathrm{d}\beta_i(m)(t)$

[9] [11] Manelli (1996) has pointed out that for sequential rationality one might as well require that this condition holds only at *almost all* information sets. In finite Bayesian games, however, sequential rationality has to hold at all information sets (even at information sets off the equilibrium path) and hence for (potentially) infinite Bayesian games it seems more natural to require sequential rationality at *all* information sets – as is done by most authors (including Manelli).

[10] For some of our results we have to require that $\sigma_i^1(m) = 1$ if $\int_T u_i(m[\sigma_0^2(t,m,1^n),$ $\sigma_1^2(m,1^n),\ldots,\sigma_n^2(m,1^n)],t)\mathrm{d}\beta_i(m)(t) > \int_T u_i(d_0,t)\mathrm{d}\beta_i(m)(t)$. This requirement implies condition (b), but for $n > 1$ it is more stringent than what is induced by sequential rationality.

for all $m \in M$ and for all $s_i \in S^i_m$ which lead to

$$t \mapsto u_i(m[\sigma^2_0(t, m, 1^n), \sigma^2_1(m, 1^n), \ldots, \sigma^2_{i-1}(m, 1^n), s_i, \sigma^2_{i+1}(m, 1^n), \ldots, \sigma^2_n(m, 1^n)], t)$$

being measurable.

This condition guarantees the optimality of agent i's action in the mechanism execution stage. The measurability-restriction is required to make sure that agent i can calculate his expected utility from playing s_i (i.e. to make sure that the integral on the right hand side of the inequality is well-defined)[11].

4. **Bayesian Consistency**
   Bayesian consistency says that interim beliefs are derived from the prior belief, P, and from the principal's strategy, σ^1_0, using Bayes' Rule (whenever possible).

The last requirement for a PBE of the contract proposal game (Bayesian Consistency) has deliberately been stated in a rather vague manner. We shall be more precise about this after having introduced *regular* strategies and beliefs in the next section.

3 Pure-Strategy Equilibrium Allocation Rules

In this section we shall analyze the mechanism design problem in terms of *allocation rules*. To do so we shall restrict ourselves to *regular* strategies and beliefs as defined below.

3.1 Regular Strategies and Beliefs

From now on we shall restrict the analysis of the contract proposal game to equilibria satisfying two *regularity* conditions. First, we shall focus on pure-strategy PBE where all agents accept the mechanism proposed by the principal, i.e. equilibria (σ, β) satisfying $\sigma^1_i(\sigma^1_0(t)) = 1 \ \forall \ i = 1, \ldots, n, \ \forall \ t \in T$. It is a trivial task for the principal to design a mechanism which yields

[11] As [6] Ellsberg (1961) and experimental studies carried out thereafter have shown, people seem to avoid ambiguous gambles if alternative *equivalent* uncertain, though unambiguous gambles exist. The measurability-restriction could now be motivated by arguing that agent i would never choose an action s_i which does not fulfill this restriction, since such a choice leads to ambiguity (about the resulting *expected* utility).

her the status quo[12], hence PBE *with acceptance* clearly form the interesting set of PBE.
Furthermore, to proceed with our analysis in the next subsections we have to use more measurability assumptions or to restrict the principal's available strategies in a suitable manner. So far we required only one measurability condition for characterizing pure-strategy equilibria of the contract proposal game; in particular, we were able to do our analysis without resorting to the σ-algebra $\mathcal{M}$. The next definition, however, specifies some measurability assumptions involving $\mathcal{M}$ or, alternatively, restricts the mechanism proposal strategy for the principal to *elementary* functions.

Definition 1 (Regular Strategies and Beliefs)
A set of strategies and beliefs $(\sigma^1, \sigma^2, \beta)$ for the contract proposal game is regular, if $\sigma_i^1(\sigma_0^1(t)) = 1$ for all $i = 1, \ldots, n$, for all $t \in T$ (acceptance), and if $(\sigma^1, \sigma^2, \beta)$ satisfies at least one of the two following conditions:

1. *Measurability*
   $\sigma_0^1 : (T, \mathcal{B}) \to (M, \mathcal{M})$ *and* $\beta_i(\cdot)(T_{\mathcal{B}}) : (M, \mathcal{M}) \to [0,1]$ *(for all* $i = 1, \ldots, n$, $\forall\ T_{\mathcal{B}} \in \mathcal{B}$*) are measurable. Furthermore,*

   $$g_i : (T \times M, \mathcal{B} \otimes \mathcal{M}) \to [-U^{max}, U^{max}]$$
   $$(t, m) \mapsto u_i(m[\sigma_0^2(t, m, 1^n), \sigma_1^2(m, 1^n), \ldots, \sigma_n^2(m, 1^n)], t)$$

   is measurable for all $i = 1, \ldots, n$. *Here* $\mathcal{B} \otimes \mathcal{M}$ *denotes the* σ*-algebra generated by* $\{T_{\mathcal{B}} \times M_{\mathcal{M}} \mid T_{\mathcal{B}} \in \mathcal{B},\ M_{\mathcal{M}} \in M_{\mathcal{M}}\}$.

2. *Elementary Mechanism Proposal Strategy*
   $\sigma_0^1 : T \to M$ *is elementary, i.e. there exists a countable index set* N, *and for each* $k \in N$ *some Borel set* T_k *and a game form* $m_k \in M$, *such that* $\{T_k\}_{k \in N}$ *forms a partition of* T, $m_k \neq m_l$ *for all* $k \neq l$, *and such that for all* $k \in N$, $t \in T_k$: $\sigma_0^1(t) = m_k$[13].

The additional measurability assumptions specified in Definition 1 seem to be fairly innocuous at first sight, and it is tempting to just make them part

[12] The principal could propose a game form which implements the status quo irrespective of the players' actions (and she could do this in such a way that in equilibrium exactly the same information is conveyed by the selection of the game form) – provided the set M does not restrict the principal's choice of mechanism too much.

[13] This definition of elementary functions as functions which assume a countable number of values and assume each value on a measurable set follows [14] Munroe (1971).

of the equilibrium requirements stated in the previous section. These assumptions, however, do depend on $\mathcal{M}$ and, hence, may turn out to be both rather incomprehensible and restrictive (if $\mathcal{M}$ has a complicated structure). In case the assumptions are considered too restrictive, we provide a second condition to make $(\sigma^1, \sigma^2, \beta)$ regular, the restriction to elementary mechanism proposal strategies. The substantive part embodied in this restriction is the countability of the index set N[14]. Though this condition is rather restrictive too, it allows for the case that T or M is countable. Furthermore, looking at elementary mechanism proposal strategies provides some kind of lower bound on what the principal can achieve in the mechanism game, and since the partition $\{T_k\}_{k\in N}$ used in Definition 1 can be chosen arbitrarily fine, one might argue that this lower bound serves as a good approximation to what the principal can achieve with arbitrary mechanism proposal strategies (at least if M is restricted appropriately). Such reasoning, however, requires the contract proposal game to be continuous in an appropriate sense[15].

For regular strategies and beliefs we can now be more specific as far as Bayesian Consistency is concerned. First, assume that $(\sigma^1, \sigma^2, \beta)$ satisfies condition (1) of Definition 1. Then Bayesian Consistency is expressed by the following condition (for $i = 1, \dots, n$):

$$\forall\ T_{\mathcal{B}} \in \mathcal{B},\ M_{\mathcal{M}} \in \mathcal{M}:\ (\sigma_0^1 \bullet P)(T_{\mathcal{B}}, M_{\mathcal{M}}) = \int_{M_{\mathcal{M}}} \beta_i(m)(T_{\mathcal{B}}) \mathrm{d}\rho(m) \quad (1)$$

[14]Instead of assuming that N is countable, it would be sufficient to assume that N has a countable subset $\widetilde{N}$ such that $\sum_{k\in\widetilde{N}} P(T_k) = 1$, i.e. that N is *essentially* countable. Furthermore, it would be easy to accommodate any situation where the choice of a mechanism by the principal reveals the principal's type (fully separating equilibria).

[15]If, for example, M is restricted to a separable metric space, $\mathcal{M}$ is the σ-algebra induced by the metric of M, and $\sigma_0^1 : (T, \mathcal{B}) \to (M, \mathcal{M})$ is measurable, then there exists a sequence $\{\sigma_k\}_{k\in\mathbb{N}}$ of elementary functions $\sigma_k : T \to M$ such that $\sigma_k \to \sigma_0^1$ uniformly on T, as can be shown by a modification of Lemma 1.1 in [5] Da Prato and Zabczyk (1992). So we can approximate σ_0^1 arbitrarily closely by elementary mechanism proposal strategies. Still, whether or not the lower bound given by elementary mechanism proposal strategies is a good approximation of what could be achieved with arbitrary measurable mechanism proposal strategies depends on the continuity-properties of the contract proposal game. The functions $u_i(\cdot, t)$ $(i = 1, \dots, n)$ being continuous in D (D a metric space), for example, may help in this respect, but this is by no means sufficient to establish continuity (in the required sense).

where for all $T_{\mathcal{B}} \in \mathcal{B}$, $M_{\mathcal{M}} \in \mathcal{M}$

$$\begin{aligned}(\sigma_0^1 \bullet P)(T_{\mathcal{B}}, M_{\mathcal{M}}) &:= \int_{T_{\mathcal{B}}} \delta_{\sigma_0^1(t)}(M_{\mathcal{M}})\mathrm{d}P(t) \\ &= \int_{T_{\mathcal{B}}} \mathbf{1}_{\sigma_0^{-1}(M_{\mathcal{M}})}(t)\mathrm{d}P(t) \\ &= P(T_{\mathcal{B}} \cap \sigma_0^{-1}(M_{\mathcal{M}}))\end{aligned}$$

(with $\sigma_0^{-1} := (\sigma_0^1)^{-1}$, δ denoting a Dirac measure, and $\mathbf{1}$ denoting a characteristic function), and the probability measure ρ on $\mathcal{M}$ is given by

$$\rho(M_{\mathcal{M}}) := (\sigma_0^1 \bullet P)(T, M_{\mathcal{M}}) = P(\sigma_0^{-1}(M_{\mathcal{M}})) = (P \circ \sigma_0^{-1})(M_{\mathcal{M}}).$$

Due to the measurability assumptions on σ_0^1 and β_i, (1) is equivalent to

$$P(T_{\mathcal{B}} \cap \sigma_0^{-1}(M_{\mathcal{M}})) = \int_{\sigma_0^{-1}(M_{\mathcal{M}})} \beta_i(\sigma_0^1(t))(T_{\mathcal{B}})\mathrm{d}P(t) \;\forall\, T_{\mathcal{B}} \in \mathcal{B}, M_{\mathcal{M}} \in \mathcal{M} \quad (2)$$

(c.f. [2] Billingsley (1995), Theorem 16.13). Equation (2) imposes for all $T_{\mathcal{B}} \in \mathcal{B}$, $M_{\mathcal{M}} \in \mathcal{M}$, a restriction on the probability of having a principal's type in the set $T_{\mathcal{B}}$ and a mechanism proposed in the set $M_{\mathcal{M}}$ in terms of the updated beliefs, β_i, and the principal's strategy for selecting mechanisms, σ_0^1. So for general Bayesian games (i.e. Bayesian games which are not necessarily finite) Bayesian consistency "puts restrictions on the behavior of beliefs at *collections* of information sets rather than at *individual* information sets" (as is done in finite Bayesian games), as [18] Perea y Monsuwe et al. (1997) have pointed out.

Note that Bayesian consistency does not put any requirements on beliefs off the equilibrium path (apart from that any updated belief has to put measure one on the set T), as can be seen from the fact that $\beta_i(m)$ does not enter equation (2) unless $m = \sigma_0^1(t)$ for some $t \in T$.

If $(\sigma^1, \sigma^2, \beta)$ does not satisfy the first two measurability conditions of Definition 1, then equation (1) may not be well-defined. But if σ_0^1 is elementary, then Bayesian updating clearly requires that

$$\beta_i(m_l)(T_{\mathcal{B}})P(T_l) = P(T_{\mathcal{B}} \cap T_l) \text{ for all } T_{\mathcal{B}} \in \mathcal{B},\; l \in N \quad (3)$$

where $\{T_k\}_{k \in N}$ is the partition of T mentioned in Definition 1. Equation (3) states that if $P(T_l) > 0$, then $\beta_i(m_l)(T_{\mathcal{B}})$ has to be the conditional probability of $T_{\mathcal{B}}$ given the set T_l (for all $T_{\mathcal{B}} \in \mathcal{B}$), i.e. $\beta_i(m_l) = \frac{P(\cdot \cap T_l)}{P(T_l)}$. Otherwise, $\beta_i(m_l)$ may be any probability measure on the set T.

If $(\sigma^1, \sigma^2, \beta)$ satisfies both conditions of Definition 1, then equation (2) and equation (3) are equivalent, provided for each $l \in N$ there exists some $M_l \in \mathcal{M}$ such that $\sigma_0^{-1}(M_l) = T_l$ (this is the case if, for example, $\{m_l\} \in \mathcal{M}$ for all $l \in N$). To see this, note that since $\{T_k\}_{k \in N}$ is a partition of T, $\{T_k \cap \sigma_0^{-1}(M_{\mathcal{M}})\}_{k \in N}$ is a partition of $\sigma_0^{-1}(M_{\mathcal{M}})$ (for any $M_{\mathcal{M}} \in \mathcal{M}$) satisfying $\sigma_0^1(t) = m_k$ for all $k \in N,\ t \in T_k \cap \sigma_0^{-1}(M_{\mathcal{M}})$. Hence, using the countability of N (and some convergence result like, for example, the monotone convergence theorem of Beppo-Levi, or Theorem 16.2 in [2] Billingsley (1995)) we can rewrite $\int_{\sigma_0^{-1}(M_{\mathcal{M}})} \beta_i(\sigma_0^1(t))(T_{\mathcal{B}})\mathrm{d}P(t)$ as

$$\begin{aligned} & \int_{\bigcup_{k \in N}(T_k \cap \sigma_0^{-1}(M_{\mathcal{M}}))} \beta_i(\sigma_0^1(t))(T_{\mathcal{B}})\mathrm{d}P(t) \\ = & \sum_{k \in N} \int_{T_k \cap \sigma_0^{-1}(M_{\mathcal{M}})} \beta_i(m_k)(T_{\mathcal{B}})\mathrm{d}P(t) \\ = & \sum_{k \in N} \beta(m_k)(T_{\mathcal{B}})P(T_k \cap \sigma_0^{-1}(M_{\mathcal{M}})). \end{aligned}$$

Therefore, equation (2) requires $\forall\ T_{\mathcal{B}} \in \mathcal{B},\ M_{\mathcal{M}} \in \mathcal{M}$

$$P(T_{\mathcal{B}} \cap \sigma_0^{-1}(M_{\mathcal{M}})) = \sum_{k \in N} \beta_i(m_k)(T_{\mathcal{B}})P(T_k \cap \sigma_0^{-1}(M_{\mathcal{M}})).$$

Now, for all $l \in N$, consider $M_l \in \mathcal{M}$ with $\sigma_0^{-1}(M_l) = T_l$. Since $P(T_k \cap \sigma_0^{-1}(M_l)) = P(T_l)$ if and only if $k = l$ (otherwise $P(T_k \cap \sigma_0^{-1}(M_l)) = 0$), equation (3) follows. On the other hand, equation (3) implies equation (2), as can easily be verified recalling that for all $M_{\mathcal{M}} \in \mathcal{M}$, $\sigma_0^{-1}(M_{\mathcal{M}}) = \bigcup_{k \in \widetilde{N}} T_k$ for a suitable subset $\widetilde{N}$ of N.

3.2 Allocation Rules

Any pure-strategy PBE for the contract proposal game implements as equilibrium outcome a certain distribution of allocations, i.e. a certain allocation $d \in D$ for any type $t \in T$. Such a social choice function $\varphi : T \to D$ shall be called an *allocation rule.* When analyzing the contract proposal game, our main interest actually concerns such allocation rules which can be implemented through perfect Bayesian equilibria, i.e. equilibrium allocation rules, rather than, for example, certain equilibrium strategies for the mechanism execution stage of the game.

The following definitions concerning allocation rules will turn out especially helpful when characterizing the set of equilibrium allocation rules for the contract proposal game.

Definition 2 (Incentive-Compatible Allocation Rules)
An allocation rule $\varphi : T \to D$ is incentive-compatible (IC), if $u_0(\varphi(t), t) \geq u_0(\varphi(\tilde{t}), t) \;\forall\; t, \tilde{t} \in T$.

Definition 3 (Individual Rationality for the Principal)
An allocation rule $\varphi : T \to D$ is individual rational for the principal (IRP), if $u_0(\varphi(t), t) \geq u_0(d_0, t) \;\forall\; t \in T$.

Definition 4 (Individual Rationality for Agent i)
An allocation rule $\varphi : T \to D$ is (ex-ante) individual rational for agent i (IRi) (for $i \in \{1, \ldots, n\}$), if $\int_T u_i(\varphi(t), t) \mathrm{d}P(t) \geq U_i^0$ (in particular, φ is such that the function $t \mapsto u_i(\varphi(t), t)$ is measurable).

If an allocation rule is both (IRP) and (IRi) for all agents i, then it shall simply be called *individual rational*. Finally, we define *strict domination* of an allocation rule.

Definition 5 (Strictly Dominated Allocation Rules)
An allocation rule $\varphi : T \to D$ is strictly dominated at $\tau \in T$ by $\psi : T \to D$, if ψ satisfies

1. *ψ is (IC),*
2. *$t \mapsto u_i(\psi(t), t)$ is measurable for all i,*
3. *$u_i(\psi(t), t) > u_i(d_0, t) \;\forall\; t \in T,\; i = 1, \ldots, n$, and*
4. *$u_0(\psi(\tau), \tau) > u_0(\varphi(\tau), \tau)$.*

We shall see later that allocation rules which are strictly dominated as defined above cannot be equilibrium allocation rules[16].

[16]Being strictly dominated is closely related to not weakly Pareto dominating the Rothschild-Stiglitz-Wilson allocation (associated with the status quo) of [13] Maskin and Tirole (1992), as will be shown in Subsection 3.3.

3.3 PBE Allocation Rules

We shall turn now to the characterization of equilibrium allocation rules for the contract proposal game. To do so, we make the convention that for an allocation rule φ, we shall write $\varphi \in M$ if and only if there exists some $m \in M$ such that $S_m^0 = T$ and $|S_m^i| = 1$ (i.e. each agent i has only one feasible action in the mechanism execution stage) and $m(t, s_1, \ldots, s_n) = \varphi(t) \ \forall \ (t, s_1, \ldots, s_n) \in T \times S_m^1 \times \ldots \times S_m^n$. Moreover, we shall say that an allocation rule φ is a PBE of the contract proposal game, if there exists some PBE (σ, β) such that $\sigma_0^1(t) = \varphi$ for all $t \in T$, and φ is accepted by all agents. Note the difference to φ *being implemented* by some PBE (σ, β) – in the former case, φ is the game form proposed by all types of principals, whereas in the later case we do not know what kind of game forms are proposed by the principal, we only know that the allocation implemented for any $t \in T$ is given by the allocation rule φ (φ being a PBE implies, of course, that φ is also the equilibrium allocation rule).
Our first theorem gives necessary conditions for an allocation rule to be implementable by a pure-strategy PBE satisfying the regularity condition of Definition 1. The proof of Theorem 1 is provided in the appendix.

Theorem 1
Suppose the status quo allocation rule ($t \mapsto d_0$) is contained in the set M. Then every pure-strategy PBE of the contract proposal game which satisfies the regularity condition implements an allocation rule φ which is incentive-compatible, individual rational for the principal, ex-ante individual rational for each agent i, and which is not strictly dominated at any $t \in T$ by any $\psi \in M$[17]. Furthermore, if $\varphi \in M$, then φ is a PBE.

The second part of Theorem 1 says that any allocation rule φ which can be implemented through a regular pure-strategy PBE is itself a PBE (provided the allocation rule is contained in the set M). Thus, instead of proposing some (complicated) game to be played, the principal might just as well announce the incentive-compatible allocation rule φ. Therefore, in our search for equilibria of the contract proposal game we can restrict ourselves to di-

[17]For the last necessary condition mentioned we have to assume that an agent accepts an allocation rule if his expected utility from the allocation rule is strictly greater than his expected utility from the status quo, or, alternatively, that $n = 1$.

rect revelation mechanisms[18]. Furthermore, if all types of principal propose φ, the principal does not reveal any of her private information at the mechanism proposal stage of the contract proposal game (any revelation of private information may be postponed to the execution stage of the contract proposal game). This is what Myerson calls the Principle of Inscrutability. So the second part of Theorem 1 establishes some revelation and inscrutability principle (the revelation part refers to the mechanism execution stage, the inscrutability part to the mechanism proposal stage). Given that this revelation and inscrutability principle holds, it is also easy to derive the necessary conditions stated in the first part of Theorem 1: Consider some direct revelation mechanism φ which in equilibrium is announced by all types of principal. Since φ is a direct revelation mechanism, it has to be an incentive-compatible allocation rule. Moreover, since the choice of φ by the principal in the mechanism selection stage of the contract proposal game does not reveal any information (Principle of Inscrutability), for φ to be accepted by an agent it has to be ex-ante individual rational for that agent (as defined in Subsection 3.2). The individual rationality constraint for the principal has to hold since, otherwise, at least one type of principal would have an incentive to implement the status quo (an option which is open to all types of principal). Similarly, strictly dominated allocation rules cannot be equilibrium allocation rules, if the dominating allocation rule is contained in the set M and hence could be implemented, as such an allocation rule would be preferred by at least one type of principal.

For their finite-type mechanism design model, [13] Maskin and Tirole (1992) have shown (in Proposition 6) that each PBE is given by an incentive-compatible allocation rule that is individual rational for the agent and that weakly Pareto dominates the *Rothschild-Stiglitz-Wilson* (RSW) allocation rule (or *allocation*, as Maskin and Tirole call it) associated with the status quo. Theorem 1 provides some generalization of this result to mechanism

[18] At first sight it may seem that we have extended the Revelation Principle, which is well-established for uninformed principal mechanism design, to our informed principal mechanism design model. But we do not prove that instead of announcing some mechanism, the principal can always announce the *associated* direct revelation mechanism. We only show that the principal can announce *some* direct revelation mechanism which will yield her the same utility. Such mechanisms are not really realistic, as [13] Maskin and Tirole (1992) have already pointed out. But they are constructs that may turn out to be very useful for theoretical analyses.

design models with more general type sets and distributions of information (given our constraint on pure-strategy PBE with "acceptance"). To see this, note that in a finite type model weakly Pareto dominating the RSW allocation rule is equivalent to being not strictly dominated (at any $t \in T$), if the third condition of Definition 5 required $u_i(\psi(t), t)$ to be greater than *or equal to* $u_i(d_0, t)$ for all $t \in T$. Moreover, with this modification any allocation rule φ which is not strictly dominated at any $t \in T$ also satisfies (IRP)[19]. The modification of Definition 5 would be feasible (i.e. could be made without changing any of the results in this paper), if each allocation specified a transfer payment plus some other decision variable, as is the case in the model of [13] Maskin and Tirole (1992). Hence, for this finite type model our necessary conditions for a PBE reduce to the same conditions as the ones derived by Maskin and Tirole.
Theorem 1 allows us to abstract from specific game forms and to just concentrate on incentive-compatible, individual rational allocation rules which are not strictly dominated. Unfortunately, it is not straight forward to prove that any allocation rule which is (IC), (IRP), (IRi) for all i, and which is not strictly dominated, is a PBE. It is easy to do so if for all $m \in M$ there exists some Bayesian Nash equilibrium of m which is independent of the beliefs of the agents. This is the case if, for example, M is restricted exogenously to allocation rules. For the case that M is not restricted in such a manner, it can be shown at least that any incentive-compatible and individual rational allocation rule can be implemented through a Bayesian equilibrium:

Theorem 2
Every allocation rule contained in the set M which is incentive-compatible, individual rational for the principal, and ex-ante individual rational for all agents can be implemented through a Bayesian equilibrium of the contract proposal game. Moreover, suppose

$$M \subseteq \{m : T \to D \mid \forall_{t \in T}\ \exists_{\tau_t^m \in T}\ \forall_{s \in T}\ u_0(m(\tau_t^m), t) \geq u_0(m(s), t) \text{ and } t \mapsto u_i(m(\tau_t^m), t) \text{ is measurable } \forall\ i > 0\}^{20}.$$

Then any allocation rule in M which is (IC), (IRP), (IRi) for all i, and which is not strictly dominated at any $t \in T$ by any $\psi : T \to D$, is a PBE.

[19] Otherwise, the allocation rule $t \mapsto d_0$ strictly dominated φ at some $t \in T$.

[20] The restrictions on M are required to ensure that optimal strategies off the equilibrium path exist for the mechanism evaluation and execution stages. All incentive-compatible allocation rules ψ with $t \mapsto u_i(\psi(t), t)$ being measurable for all i satisfy these requirements.

4 Equilibrium Refinements

The set of all potential perfect Bayesian equilibrium allocation rules for the contract proposal game identified in Theorem 1, i.e. the set of all allocation rules which are (IC), (IRP), (IRi) for all i, and which are not strictly dominated, may be fairly large, due to the fact that beliefs in PBE are arbitrary off the equilibrium path – as pointed out earlier. Therefore, in this section we shall present various equilibrium refinements to reduce the set of possible equilibrium allocation rules. In order to facilitate the development of appropriate refinement criteria, in the first subsection we shall look at a game which is closely related to the contract proposal game, the *mechanism signaling game*.

4.1 The Mechanism Signaling Game

One class of economic models for which equilibrium refinements have been studied extensively is the class of signaling games. Our original contract proposal game is more complex than a signaling game (due to the mechanism execution stage of the game), but due to Theorem 1 we could require the principal to propose some incentive-compatible allocation rule φ[21], and then we could eliminate the mechanism execution stage. The payoff for each player is then given by $u_i(\varphi(t), t)$, if φ is accepted. Thus, the contract proposal game reduces to a signaling game with one sender and n receivers. Moreover, again by invoking Theorem 1, we can restrict ourselves to allocation rules φ such that $t \mapsto u_i(\varphi(t), t)$ is measurable for all i.
So define

$$M^* := \{m : T \to D \mid m \ (IC),\ t \mapsto u_i(m(t), t) \text{ is measurable } \forall\ i = 1, \dots, n\},$$

and let us consider the following *mechanism signaling game*:

[21]Focusing on pure and regular strategies (and beliefs), we know that each PBE implements some incentive-compatible allocation rule φ, which itself is a PBE (provided $\varphi \in M$). Hence, if the principal would want to announce some other game form in M, the agents might approach the principal and require her to announce some incentive-compatible allocation rule. Thus the restriction to incentive-compatible mechanisms does not seem to be severe. However, this restriction is not completely innocuous: Though restricting the principal's mechanism to incentive-compatible allocation rules does not change much on the equilibrium path, it restricts the principal from sending certain messages off the equilibrium path.

1. Nature chooses the principal's type $t \in T$.
2. The principal (the sender) observes her type and then announces an allocation rule $m \in M^*$ (the signal or message).
3. The n agents (receivers) observe the signal m (but not the sender's type), update their prior belief about the sender's type, and decide whether to accept or to reject m.
4. If all agents have accepted m, player i receives the payoff $u_i(m(t), t)$, otherwise player i receives $u_i(d_0, t)$ $(i = 0, \ldots, n)$.

Lemma 1 characterizes the set of pure-strategy perfect Bayesian *pooling* equilibria for this mechanism signaling game. The only minor difficulty in the proof of Lemma 1, which is given in the appendix, concerns the behavior off the equilibrium path (this is where the property of not being strictly dominated comes into play).

Lemma 1
Assume that an agent accepts an allocation rule if his expected utility from the allocation rule is greater than his expected utility from the status quo. Then the set of pure-strategy perfect Bayesian pooling *equilibria for the mechanism signaling game involving play of 'accept' by all agents is given by*

$$E := \{m : T \to D \quad | \quad m \ (IC),\ (IRP),\ (IRi)\ \forall\ i = 1, \ldots, n;\ m \textit{ is not strictly dominated at any } t \in T \textit{ by any } \widetilde{m} \in M^*\}.$$

The set E is precisely the set of allocation rules which we have identified as potential equilibrium allocation rules for the contract proposal game (given our restriction on pure and regular strategies). This link between equilibria of the contract proposal game and pooling equilibria of the mechanism signaling game once again gives some justification why we may focus on the mechanism signaling game – when analyzing the contract proposal game we can actually focus on the mechanism signaling game due to the revelation part of Theorem 1, and due to the Principle of Inscrutability we can focus on pooling equilibria of this game. Hence, we proceed by analyzing the mechanism signaling game and by deriving appropriate refinements for pooling equilibria of this game.

4.2 Intuitive Criterion

Refinements of the perfect Bayesian equilibrium concept try to put reasonable restrictions on beliefs off the equilibrium path (as said earlier, the main problem with the concept of PBE is that in a PBE no restrictions are made on such beliefs). The commonly used concept of sequential equilibria of [9] Kreps and Wilson (1982) is, unfortunately, not of much help here, since for a signaling game the set of sequential equilibria and the set of perfect Bayesian equilibria (usually) coincide[22].

[4] Cho and Kreps (1987) give an overview over some equilibrium refinements for finite signaling games and develop some new refinement criteria. Possibly the weakest equilibrium refinement is based on the idea that equilibrium beliefs should not assign positive probability to types t after observing out of equilibrium signals which constitute a dominated action for such a type t (i.e. which are certain to be inferior to some other signal for type t). Cho and Kreps extend this idea to postulating that we should not expect that some type of sender sends a signal which is certain to result in an inferior outcome for this sender compared to what this sender receives in equilibrium.

From this idea of *equilibrium domination*, Cho and Kreps develop a refinement called the *intuitive criterion.* To eliminate an *unreasonable* PBE it requires that there exists some type $\tilde{t}$ of sender who would like to deviate from the equilibrium signal m to some other signal $\widetilde{m}$, provided that all receivers will deduce from the choice of signal $\widetilde{m}$ that the sender does not belong to those types for which $\widetilde{m}$ constitutes a signal which is equilibrium dominated (and that the receivers update their prior beliefs and choose a response to $\widetilde{m}$ accordingly). So the intuitive criterion puts some restrictions on reasonable beliefs off the equilibrium path – reasonable beliefs must not put positive weight on types of sender for which the signal is equilibrium dominated, and it requires the sender to take this into account when choosing optimal actions[23].

[22] [7] Fudenberg and Tirole (1991) have demonstrated that the set of PBE and the set of sequential equilibria coincide for two-period games of incomplete information with finite type sets. [18] Perea y Monsuwe et al. (1997) have provided some extension of this result to infinite signaling games.

[23] If one accepts the idea that defections from the equilibrium path represent attempts of players to do better than on the equilibrium path, then the intuitive criterion is an obvious refinement concept. Though it has proven very helpful for the analysis of signaling

Let us apply the intuitive criterion to the mechanism signaling game. So suppose that φ is some incentive-compatible, individual rational allocation rule, which is expected to be played in some pooling equilibrium of the mechanism signaling game. Now consider some other message $\psi \in M^*$ (i.e. some signal off the equilibrium path), and define the set $T(\psi) := \{t \in T \mid u_0(\psi(t), t) > u_0(\varphi(t), t)\}$. This is the set of all types of principal which strictly prefer ψ over the equilibrium allocation rule φ. Hence we can deduce by equilibrium domination that $t \in T(\psi)$ if an incentive-compatible allocation rule $\psi \neq \varphi$ has been proposed by the principal (provided $T(\psi) \neq \emptyset$)[24]. For φ to be an equilibrium of the mechanism signaling game, the intuitive criterion then requires that there does not exist any incentive-compatible allocation rule $\psi : T \to D$ (satisfying that $t \mapsto u_i(\psi(t), t)$ is measurable for all agents i) such that $T(\psi) \neq \emptyset$ and such that accepting ψ is the only best response of the agents given that $t \in T(\psi)$. This leads to the following solution concept for the mechanism signaling game:

Definition 6 (Intuitive Solution)
An allocation rule $\varphi : T \to D$ is an intuitive solution, if it satisfies the following conditions:

1. *φ is (IC), (IRP), (IRi) $\forall\ i = 1, \ldots, n$.*
2. *Suppose $\psi : T \to D$ is (IC), $\psi \in M$, $t \mapsto u_i(\varphi(t), t)$ is measurable for all i, and $S = \{t \in T \mid u_0(\psi(t), t) > u_0(\varphi(t), t)\} \neq \emptyset$. Then there exists some $j \in \{1, \ldots, n\}$ and some $t \in S$[25] such that $u_j(\psi(t), t) \leq u_j(d_0, t)$.*

games, the intuitive criterion, like *forward induction* in general, is not met with unanimous approval. It has been criticized by, among others, [10] Mailath et al. (1993) by pointing out that the refinement is based on agents inferring some information from other players' deviations from the equilibrium path, but nothing is inferred from players *not* deviating from equilibrium play (this leads to the question of whether the equilibrium outcome can really be viewed as the benchmark against which deviations have to be evaluated).

[24] Here we deviate slightly from the standard argument of equilibrium domination, as we assume that the principal deviates from equilibrium play only if she *strictly* benefits from such a deviation. This variation of standard equilibrium domination, which is not essential for the analysis to follow, can be justified by observing that deviating from equilibrium play may be more risky in some sense (and why should the principal take such risks if she does not expect to strictly benefit from taking the risk).

[25] If we had applied the intuitive criterion in the original sense of [4] Cho and Kreps (1987), then we would have had to find some $t \in \{t \in T \mid u_0(\psi(t), t) \geq u_0(\varphi(t), t)\}$ to establish φ as an intuitive solution – c.f. footnote No. 24.

Note that any intuitive solution is not strictly dominated at any $t \in T$ by any $\psi \in M$ (otherwise condition (2) would be violated). Thus, any intuitive solution is contained in the set E specified in Lemma 1.
If we assume that agents accept an allocation rule whenever it yields them an expected utility greater than their expected utility from the status quo, then with the help of (the proof of) Lemma 1 it can easily be shown that any pure-strategy pooling PBE of the mechanism signaling game involving play of 'accept' by all agents and satisfying the intuitive criterion implements an intuitive solution (and vice versa, see footnote No. 33).

4.3 Perfect Solution

Though the intuitive criterion is indeed fairly intuitive, this equilibrium refinement may turn out to be relatively weak. This is so because after it has been deduced that certain signals cannot have been sent by certain senders (as long as these senders are rational and they understand that the receivers are rational as well), the intuitive criterion does put some restriction on the *support* of reasonable updated beliefs, but it does not put any restriction on the *shape* of the updated distribution function. Therefore, let us look at the following procedure for forming reasonable beliefs off the equilibrium path: Suppose that the receivers have deduced from equilibrium domination that the sender's type has to be in the set $\widetilde{T} \subseteq T$ after observing some signal $\widetilde{m}$ off the equilibrium path, and assume that no type $t \in \widetilde{T}$ can be excluded from having sent $\widetilde{m}$ (more precisely, assume that all types $t \in \widetilde{T}$ are expected to send the signal $\widetilde{m}$ given that $\widetilde{m}$ has been sent). Then the receivers should derive their updated beliefs by evaluating conditional probabilities given the set $\widetilde{T}$ (Bayesian updating conditional on $\widetilde{T}$). Thus, in our model, if $\widetilde{T} \in \mathcal{B}$ and $P(\widetilde{T}) > 0$, then the updated belief $\beta_i(\widetilde{m})$ should be given by $\beta_i(\widetilde{m})(T_{\mathcal{B}}) = P(T_{\mathcal{B}} \cap \widetilde{T})/P(\widetilde{T})$ for all $T_{\mathcal{B}} \in \mathcal{B}$. Hence, like with Bayesian updating on the equilibrium path, beliefs off the equilibrium path are now (essentially) completely determined. With this updating procedure we have eliminated any freedom in the assignment of beliefs off the equilibrium path (which may have served to support unreasonable equilibria). We shall call such beliefs *perfect* because these beliefs are basically the beliefs which [8] Grossman and Perry (1986) proposed for their perfect sequential equilibrium.

Using perfect beliefs off the equilibrium path is, of course, not as compelling as Bayesian updating on the equilibrium path. Bayesian updating can be justified by noting that such updates are the only updates which are actually confirmed in the long run (i.e. by repeated equilibrium play). For the procedure of using perfect beliefs off the equilibrium path this is clearly not the case. But if one accepts this idea of updating beliefs off the equilibrium path, then this leads to a much stronger equilibrium refinement as compared to the intuitive criterion.

Let us look at the implications of using perfect beliefs for the mechanism signaling game. So, again, suppose that φ is some incentive-compatible, individual rational allocation rule, which is expected to be played in some pooling equilibrium of the mechanism signaling game, and consider some other message $\psi \in M^*$. As said in the previous subsection, we can now deduce by equilibrium domination that $t \in T(\psi) = \{s \in T \mid u_0(\psi(s), s) > u_0(\varphi(s), s)\}$ (provided $T(\psi) \neq \emptyset$). Moreover, suppose that given that the allocation rule ψ has been chosen, all agents assume that all types of principal in the set $T(\psi)$ would have chosen ψ[26]. Thus, φ fails to pass the equilibrium refinement test if $P(T(\psi)) > 0$ and $\int_{T(\psi)} u_i(\psi(t), t)\mathrm{d}\beta_i(\psi)(t) > \int_{T(\psi)} u_i(d_0, t)\mathrm{d}\beta_i(\psi)(t)\ \forall\ i = 1, \ldots, n$, where $\beta_i(\psi)$ denotes the prior probabilities conditional on the set $T(\psi)$. If such an allocation rule ψ exists, then all types of principal in the set $T(\psi)$ could propose ψ rather than φ, anticipating that ψ will be accepted by all agents (as agents take the signal ψ as evidence that $t \in T(\psi)$, and hence their expected utility from ψ is greater than their expected utility from the status quo (using perfect beliefs to evaluate expected utilities)), because ψ gives all types $t \in T(\psi)$ a strictly greater utility than what they receive in equilibrium. Therefore, φ cannot be an equilibrium.

It remains to specify precisely perfect beliefs off the equilibrium path. As said earlier, if $T(\psi) \in \mathcal{B}$ and $P(T(\psi)) > 0$ for some ψ off the equilibrium path, then the updated belief $\beta_i(\psi)$ is given by $\beta_i(\psi)(T_{\mathcal{B}}) = P(T_{\mathcal{B}} \cap T(\psi))/P(T(\psi)$ for all $T_{\mathcal{B}} \in \mathcal{B}$. Though $T(\psi) \in \mathcal{B}$ and $P(T(\psi)) > 0$ for

[26] This is, of course, a fairly restrictive assumption about the way agents interpret signals off the equilibrium path (c.f. [10] Mailath et al. (1993), pp. 246–248). The assumption would be appropriate if agents do not figure out or are not able to figure out which of the types in the set $T(\psi)$ are more likely, and which are less likely to send the signal ψ.

$\psi \in M^*$ is satisfied for some mechanism design problems[27], let us now specify how to proceed if these conditions do not hold.
To do so, set $S := T(\psi)$. In general, S is not necessarily a Borel set[28], so choose some $\widehat{S} \in \operatorname{argmax}\{P(\widetilde{S}) \mid S \supseteq \widetilde{S} \in \mathcal{B}\}$ (clearly, the set $\{P(\widetilde{S}) \mid S \supseteq \widetilde{S} \in \mathcal{B}\}$ assumes its supremum, and the set $\widehat{S}$ is unique up to a Borel set of measure zero). If $S \in \mathcal{B}$, then we can choose $\widehat{S} = S$, of course. If $P(\widehat{S}) > 0$, then we proceed by standard updating (as spelled out above). If, however, $P(\widehat{S}) = 0$, then we shall just require that the support of the updated belief is a subset of S (so in this case we adopt the intuitive criterion) – unless $S = \emptyset$, where we only require that the updated belief has full support on T[29].
Having specified how reasonable beliefs off the equilibrium path may look like, we are now in a position to present the following equilibrium refinement for the mechanism signaling game.

Definition 7 (Perfect Solution)
An allocation rule $\varphi : T \to D$ is a perfect solution[30]*, if it satisfies the following conditions:*

1. *φ is (IC), (IRP), and (IRi) $\forall\ i = 1, \ldots, n$.*
2. *Suppose $\psi : T \to D$ is (IC), $\psi \in M$, $t \mapsto u_i(\varphi(t), t)$ is measurable for all i, and $S = \{t \in T \mid u_0(\psi(t), t) > u_0(\varphi(t), t)\} \neq \emptyset$.*
   Let $\widehat{S} \in \operatorname{argmax}\{P(\widetilde{S}) \mid S \supseteq \widetilde{S} \in \mathcal{B}\}$.

[27] The conditions are satisfied if T is discrete. Moreover, they are satisfied if, for example, $T = [\underline{t}, \overline{t}]$ and u_0 is given by $u_0(\phi(t), t) = \phi_1(t) + \widetilde{u}_0(\phi_2(t))t$ ($\phi = (\phi_1, \phi_2)$; ϕ_1 typically denotes some transfer payment), where $\widetilde{u}_0$ is bounded on $\phi_2(T)$. Then we know from standard mechanism design theory that $t \mapsto u_0(\phi(t), t)$ is continuous (even Lipschitz continuous) for any incentive-compatible ϕ. The continuity of these functions ensures that $T(\psi)$ is an open set and hence $P(T(\psi)) > 0$ for any strictly positive probability measure P on T (i.e. any P which puts positive weight on every non-empty, open subset of T).

[28] S is a Borel set if $t \mapsto u_0(\varphi(t), t)$ and $t \mapsto u_0(\psi(t), t)$ are measurable. Thus, we can evade the *ad-hoc* assumptions which are to follow now by requiring that the allocation rule φ to be proposed by the principal leads to $t \mapsto u_i(\varphi(t), t)$ being measurable for all $i = \mathbf{0}, 1, \ldots, n$.

[29] If $T(\psi) = \emptyset$ then ψ is an allocation rule which is not preferred to the equilibrium by any type of principal.

[30] In how far the solution can really be considered *perfect* depends on the willingness to accept the assumptions made on the agents' responses to actions off the equilibrium path – c.f. footnote No. 26.

If $P(\widehat{S}) > 0$, then there exists some $j \in \{1, \ldots, n\}$ such that

$$\int_{\widehat{S}} u_j(\psi(t), t) \mathrm{d}P(t) \leq \int_{\widehat{S}} u_j(d_0, t) \mathrm{d}P(t)^{31}.$$

If $P(\widehat{S}) = 0$, then there exists some $j \in \{1, \ldots, n\}$ and some $t \in S$ such that $u_j(\psi(t), t) \leq u_j(d_0, t)$.

Intuitively it should be clear that any pooling equilibrium of the mechanism signaling game has to be a perfect solution as defined above if agents use perfect beliefs off the equilibrium path: Any pooling equilibrium has to fulfill (IRP) and (IRi) for all $i = 1, \ldots, n$, otherwise the equilibrium allocation rule would be rejected by at least one agent or not announced by at least one type of principal. Moreover, if the second condition of Definition 7 is not satisfied, then the potential equilibrium cannot be "in equilibrium", because then some type of principal would do better by announcing some other incentive-compatible allocation rule which would be accepted by all agents (as long as their beliefs off the equilibrium path are perfect).
On the other hand, any perfect solution is a perfect Bayesian equilibrium of the mechanism signaling game with perfect beliefs off the equilibrium path: If an allocation rule φ satisfying conditions (1) and (2) of Definition 7 is announced by the principal, all agents will accept φ due to (IRi). Furthermore, if the principal announced any other incentive-compatible allocation rule ψ with which she could do potentially better (i.e. $S \neq \emptyset$ in (2) of Definition 7), then this allocation rule could be rejected by at least one agent if all agents take the announcement of ψ as evidence that $t \in S$ and update their beliefs accordingly (c.f. (2) of Definition 7). The next theorem provides the verification of the intuition above.

Theorem 3
Every perfect solution is a PBE of the mechanism signaling game with perfect beliefs off the equilibrium path.
Moreover, assume that an agent accepts an allocation rule if his expected utility from the allocation rule is greater than his expected utility from the status quo. Then, conversely, every pure-strategy perfect Bayesian pooling equilibrium of the mechanism signaling game with perfect beliefs off the equilibrium path and with "acceptance" implements a perfect solution.

[31] Since $\widehat{S}$ is unique up to a Borel set of measure zero, the two integrals do not depend on the choice of $\widehat{S} \in \operatorname{argmax}\{P(\widetilde{S}) \mid S \supseteq \widetilde{S} \in \mathcal{B}\}$.

Furthermore, Theorem 4 shows that any regular pure-strategy PBE of the *contract proposal* game with perfect beliefs off the equilibrium path implements as equilibrium allocation rule a perfect solution. Unfortunately, as with Theorem 1, the converse statement is not easy to prove, since the equilibrium implemented in the mechanism execution stage by some mechanism off the equilibrium path in general[32] does depend on the beliefs off the equilibrium path. Thus *manipulating* these beliefs in such a way as to guarantee that no type of principal would like to deviate from equilibrium play, we might alter the equilibrium played in the mechanism execution stage. The proofs of Theorem 3 and Theorem 4 are to be found in the appendix[33].

Theorem 4
Suppose the status quo allocation rule ($t \mapsto d_0$) is contained in the set M, and assume that an agent accepts an allocation rule if his expected utility from the allocation rule is greater than his expected utility from the status quo.
Then every regular pure-strategy perfect Bayesian equilibrium of the contract proposal game with perfect beliefs off the equilibrium path implements as equilibrium allocation rule a perfect solution.

The last two theorems characterize the set of allocation rules we may expect as outcomes for the mechanism signaling game and the contract proposal game. This set does indeed constitute a further equilibrium refinement, as any perfect solution is clearly also an intuitive solution.
A natural question to ask at this point concerns the existence of perfect solutions. Unfortunately, as [22] Tisljar (2002) has shown, such solutions need not exist, even if all utility functions in the model are well behaved[34].

4.4 Undominated Perfect Solution

We shall now present a further refinement criterion inspired by [16] Myerson (1983). As will be pointed out later, the validity of this criterion (like the

[32]If $M = M^*$, then this is not the case, hence in Theorem 3 we were able to prove the converse statement.

[33]Results analogous to Theorem 3 and Theorem 4 hold for the intuitive solution. The proofs of these results are very similar to the proofs of Theorem 3 and Theorem 4 (and actually slightly easier), hence these corresponding results are not proven separately.

[34][8] Grossman and Perry (1986) demonstrate that perfect sequential equilibria may not exist. But they propose that if a perfect sequential equilibrium exists, then it should be used as refined equilibrium concept.

validity of the refinement criterion presented in the previous subsection) is debatable; nevertheless, it shall be presented here, as it is required for comparisons in the following section. Myerson's ideas suggest that we should not expect any allocation rule ψ to evolve as outcome of the contract proposal game, if there exists a perfect solution φ which weakly dominates ψ, i.e. $u_0(\varphi(t), t) \geq u_0(\psi(t), t) \ \forall \ t \in T$ and $\exists_{t \in T} \ u_0(\varphi(t), t) > u_0(\psi(t), t)$[35]. Since ψ is a perfect solution, φ may be rejected by at least one agent if the announcement of φ is to be taken by the agents as evidence that the principal's type is in the set $\{t \in T \mid u_0(\varphi(t), t) > u_0(\psi(t), t)\}$. If the principal, however, can convince the agents not to make such an inference pointing out that *all* of her types weakly prefer φ over ψ, then φ remains implementable, and it may be reasonable to assume that the principal does not announce ψ. If we accept this argument, we can limit the set of outcomes of the contract proposal game to *undominated* perfect solutions as defined below.

Definition 8 (Undominated Perfect Solution)
An allocation rule φ is an undominated perfect solution if it is a perfect solution and if it is not dominated by any other perfect solution, i.e. if there does not exist any perfect solution ψ such that $u_0(\psi(t), t) \geq u_0(\varphi(t), t)$ for all $t \in T$ and $u_0(\psi(t), t) > u_0(\varphi(t), t)$ for some $t \in T$.

It shall be pointed out here once more that the argument propagating the undominated (rather than any other) perfect solution relies on the principal being able to convince the agents not to infer anything from exchanging a dominated by some undominated perfect solution. When discussing the concept of (strict) equilibrium domination, we argued that some type of principal might bother deviating from some equilibrium play only if she strictly benefited from doing so. Thus the argument for the undominated perfect solution may break down[36], so that one might expect dominated perfect solutions to be as likely to evolve as equilibrium play as any other perfect solution. Nevertheless, the concept of undominated perfect solutions does have some appeal once one starts to think about efficiency in the equilibrium selection process (efficiency from the point of view of the principal).

[35] A similar argument could be used to eliminate any PBE which is dominated by some other PBE.

[36] Myerson actually gives the argument only to justify that any solution φ with $u_0(\varphi(t), t) < u_0(\psi(t), t)$ for *all* $t \in T$ is unreasonable in his model, but then he goes on discarding any solution which is *weakly* dominated in the sense spelled out above.

5 Strong Solutions

In the previous section we have presented the perfect solution as solution concept for the contract proposal game, and we have argued that the set of solutions to be expected as outcome of this game may be restricted further to undominated perfect solutions. For the case that even the undominated perfect solution is not unique, we shall now present an equilibrium concept proposed by [16] Myerson (1983) which identifies a solution which is essentially unique – in case it exists. Myerson's insights do not directly apply to our model (as he considers a finite type model with adverse selection and moral hazard), but the following ideas can be used for our model as well. Myerson suggests to look at *strong solutions*, which are defined in our model as follows:

Definition 9 (Feasible Allocation Rules)
An allocation rule $\varphi : T \to D$ is feasible, if it is (IC) and (IRi) for all $i = 1, \dots, n$. The set of all feasible allocation rules is denoted by $\mathcal{F}$.

Definition 10 (Undominated Allocation Rules)
An allocation rule $\varphi \in \mathcal{F}$ is dominated, if there exists some other allocation rule $\psi \in \mathcal{F}$ such that $u_0(\varphi(t), t) \leq u_0(\psi(t), t)$ for all $t \in T$ and $u_0(\varphi(t), t) < u_0(\psi(t), t)$ for some $t \in T$.
An allocation rule φ is undominated, if $\varphi \in \mathcal{F}$ and if there does not exist any $\psi \in \mathcal{F}$ which dominates φ.

Definition 11 (Safe Allocation Rules)
An incentive-compatible allocation rule φ is safe, if $u_i(\varphi(t), t) \geq u_i(d_0, t)$ for all $i \in \{0, \dots, n\}$, $t \in T$.

Definition 12 (Strong Solution)
An allocation rule $\varphi : T \to D$ is a strong solution, if φ is undominated and safe.

From Theorem 1 we know that in our search for pure-strategy equilibria of the contract proposal game (satisfying the regularity restriction) we can restrict ourselves without loss of generality to incentive-compatible allocation rules. Furthermore, for such rules to be accepted by all agents we need as additional minimum requirement that $\int_T u_i(\varphi(t), t) \mathrm{d}P(t) \geq U_i^0 \; \forall \; i \in \{1, \dots, n\}$. Hence we can focus our attention on feasible allocation rules.

Moreover, like in the previous section, it may be argued that among all feasible allocation rules we can restrict ourselves to undominated allocation rules[37].

A safe allocation rule is guaranteed to be implementable by the principal irrespective of what the agents' beliefs about the principal's type look like, in particular irrespective of how the prior belief, given by the probability measure P, is updated. Safe allocation rules may be dominated, and strong solutions may fail to exist. But the next theorem shows (under some regularity condition) that if a strong solution exists, then it is an undominated perfect solution and it is essentially unique (that is unique from the point of view of the principal's utilities). The proof of Theorem 5 is again provided in the appendix.

Theorem 5

Let the contract proposal game be such that

$$\{\varphi : T \to D \mid \varphi \text{ IC and } t \mapsto u_i(\varphi(t), t) \text{ is measurable for all } i > 0\}$$
$$= \{\varphi : T \to D \mid \varphi \text{ IC and } t \mapsto u_0(\varphi(t), t) \text{ is measurable}\}.$$

Suppose φ is a strong solution. Then φ is an undominated perfect solution of the contract proposal game. Moreover, if $\widetilde{\varphi}$ is some other strong solution, then $u_0(\varphi(t), t) = u_0(\widetilde{\varphi}(t), t)$ $\forall\, t \in T$.

So the strong solution as equilibrium concept has the advantage that under the condition stated[38] strong solutions are essentially unique and that they are undominated perfect solutions (and therefore PBE of the mechanism signaling game with perfect beliefs off the equilibrium path). But besides the fact that strong solutions may not exist, it is not so clear why the outcome of the contract proposal game should necessarily be the strong solution rather than any other (undominated) perfect solution.

[37]Though we have provided some motivation for this selection criterion, we have also pointed out the caveat of this approach, which equally applies to the current section.

[38]The measurability of the function $t \mapsto u_0(\varphi(t), t)$ is often sufficient to ensure the measurability of the functions $t \mapsto u_i(\varphi(t), t)$ (and vice versa), especially if there is just one agent (this is the case if, for example, the principal's and the agent's utilities sum up to a measurable function). Moreover, if $T = [\underline{t}, \overline{t}]$ and u_0 is given by $u_0(\varphi(t), t) = \varphi_1(t) + \widetilde{u}_0(\varphi_2(t))t$ ($\varphi = (\varphi_1, \varphi_2)$), where $\widetilde{u}_0$ is bounded on $\varphi_2(T)$, then $t \mapsto u_0(\varphi(t), t)$ is continuous (c.f. footnote No. 27), and $t \mapsto \widetilde{u}_0(\varphi_2(t))$ is increasing. If $\widetilde{u}_0$ is strictly increasing, then φ_2 is an increasing and φ_1 a decreasing function (provided $\underline{t} \geq 0$), which in turn may be sufficient to ensure that for all i the function $t \mapsto u_i(\varphi(t), t)$ is measurable (under suitable conditions for u_i).

6 Conclusion

We have extended the analysis of the finite type informed principal mechanism design model of [13] Maskin and Tirole (1992) to a model where the (one-dimensional) private information of the principal may be distributed according to any general probability measure. Regular pure-strategy perfect Bayesian equilibria of such a model implement allocation rules which are incentive-compatible, individual rational for the principal, ex-ante individual rational for all agents, and not strictly dominated. We have applied to our model the intuitive criterion as an equilibrium refinement. Since the intuitive criterion is a relatively weak refinement criterion, we have provided an equilibrium refinement related to the perfect sequential equilibrium of [8] Grossman and Perry (1986). This refinement is based on fairly restrictive assumptions about how agents update their beliefs off the equilibrium path, but it allows for a relatively strong equilibrium refinement criterion.
The theory presented in the paper may serve to solve common-value mechanism design problems for an informed principal. It provides a framework for the analysis of continuous-type mechanism design models, and assuming that types are continuously distributed, we can employ the differential representation of incentive-compatibility constraints for the principal (which is, of course, not the case for discrete type models), which in turn may help to characterize the set of direct revelation mechanisms. This enhancement is bought, however, by the restriction to pure-strategy equilibria satisfying some regularity condition (though the restriction to regular equilibria is redundant if the set of feasible mechanisms is countable). Further research might look at extensions of the model to mechanism design problems for informed principals with bilateral incomplete information or to informed principal problems with moral hazard.

Appendix – Proofs

Proof of Theorem 1
Suppose (σ, β) is a regular pure-strategy PBE. Hence (σ, β) is given by some functions $\sigma_0^1 : T \to M$, $\sigma_0^2 : T \times M \times A \to \bigcup_{S^0 \in \mathcal{S}^0} S^0$, $\sigma_i^1 : M \to \{0, 1\}$, $\sigma_i^2 : M \times A \to \bigcup_{S^i \in \mathcal{S}^i} S^i$, and $\beta_i : M \to \mathcal{P}(T)$ $(i = 1, \ldots, n)$. Since $\sigma_0^1(t)$ is accepted by all agents for all $t \in T$ (due to the first part of the regularity

condition), (σ, β) implements the allocation rule φ given by

$$\begin{aligned} \varphi: \quad T &\to \quad D \\ t &\mapsto \quad \sigma_0^1(t)[\sigma_0^2(t, \sigma_0^1(t), 1^n), \sigma_1^2(\sigma_0^1(t), 1^n), \ldots, \sigma_n^2(\sigma_0^1(t), 1^n)]. \end{aligned}$$

We shall prove now that φ is (IC), (IRP), (IRi) for all i, and not strictly dominated at any $t \in T$ by any $\psi \in M$.

φ is (IC)

$m := \sigma_0^1(\tilde{t}) \in M$ is accepted by all agents for all $\tilde{t} \in T$. Thus, since (σ, β) is sequential rational for the principal, we know that for all $t \in T$

$$\begin{aligned} u_0(\varphi(t), t) &= u_0(\sigma_0^1(t)[\sigma_0^2(t, \sigma_0^1(t), 1^n), \sigma_1^2(\sigma_0^1(t), 1^n), \ldots, \sigma_n^2(\sigma_0^1(t), 1^n)], t) \\ &\geq u_0(m[\sigma_0^2(t, m, 1^n), \sigma_1^2(m, 1^n), \ldots, \sigma_n^2(m, 1^n)], t) \\ &\geq u_0(m[\sigma_0^2(\tilde{t}, m, 1^n), \sigma_1^2(m, 1^n), \ldots, \sigma_n^2(m, 1^n)], t) \\ &\quad = u_0(\varphi(\tilde{t}), t), \end{aligned}$$

where the first inequality is due to part (a), and the second inequality is due part (b) of the requirement for (σ, β) being sequential rational for the principal (note $\sigma_0^2(\tilde{t}, m, 1^n) \in S^0_{\sigma_0^1(\tilde{t})}$). Hence, φ is (IC).

φ is (IRP)

Since $(t \mapsto d_0) \in M$, the sequential rationality of (σ, β) for the principal yields

$$\begin{aligned} u_0(\varphi(t), t) &= u_0(\sigma_0^1(t)[\sigma_0^2(t, \sigma_0^1(t), 1^n), \sigma_1^2(\sigma_0^1(t), 1^n), \ldots, \sigma_n^2(\sigma_0^1(t), 1^n)], t) \\ &\geq u_0(d_0, t) \end{aligned}$$

for all $t \in T$ (no matter whether or not $t \mapsto d_0$ is accepted by all agents), i.e. φ is (IRP).

φ is (IRi) for all agents i

First suppose (σ, β) satisfies the first condition of the regularity definition, i.e. $\sigma_0^1 : (T, \mathcal{B}) \to (M, \mathcal{M})$, $\beta_i(\cdot)(T_{\mathcal{B}}) : (M, \mathcal{M}) \to [0, 1]$ (for all $i = 1, \ldots, n$, $\forall\ T_{\mathcal{B}} \in \mathcal{B}$) and

$$\begin{aligned} g_i : (T \times M, \mathcal{B} \otimes \mathcal{M}) &\to \quad [-U^{max}, U^{max}] \\ (t, m) &\mapsto \quad u_i(m[\sigma_0^2(t, m, 1^n), \sigma_1^2(m, 1^n), \ldots, \sigma_n^2(m, 1^n)], t) \end{aligned}$$

are measurable. Fix some $i \in \{1, \ldots, n\}$. The idea of the following proof showing that φ satisfies (IRi) is to use iterated expectations:

Since P is a probability measure on $(T, \mathcal{B})$, $\delta_{\sigma_0^1(t)}$ is a probability measure on $(M, \mathcal{M})$ for all $t \in T$, and $\delta_{\sigma_0^1(\cdot)}(M_{\mathcal{M}}) : (T, \mathcal{B}) \to [0, 1]$ is measurable for all $M_{\mathcal{M}} \in \mathcal{M}$ (since σ_0^1 is measurable), it follows by Ex. 18.20 in [2] Billingsley (1995) for all measurable $f : (T \times M, \mathcal{B} \otimes \mathcal{M}) \to [0, \infty]$ that $\int_M f(\cdot, m) \mathrm{d}\delta_{\sigma_0^1(\cdot)}(m) : (T, \mathcal{B}) \to [0, \infty]$ is measurable and

$$\int_{T \times M} f(t, m) \mathrm{d}\pi(t, m) = \int_T \int_M f(t, m) \mathrm{d}\delta_{\sigma_0^1(t)}(m) \mathrm{d}P(t),$$

with the probability measure π being given by

$$\pi(H) = \int_T \delta_{\sigma_0^1(t)}(\{m \in M \mid (t, m) \in H\}) \mathrm{d}P(t) \quad \forall\, H \in \mathcal{B} \otimes \mathcal{M}.$$

Moreover, $\rho = P \circ \sigma_0^{-1}$ is a probability measure on $(M, \mathcal{M})$, $\beta_i(m)$ is a probability measure on $(T, \mathcal{B})$ for all $m \in M$, and $\beta_i(\cdot)(T_{\mathcal{B}}) : (M, \mathcal{M}) \to [0, 1]$ is measurable for all $T_{\mathcal{B}} \in \mathcal{B}$. Thus, again by Ex. 18.20 in [2] Billingsley (1995), it follows for all measurable $f : (T \times M, \mathcal{B} \otimes \mathcal{M}) \to [0, \infty]$ that $\int_T f(t, \cdot) \mathrm{d}\beta_i(\cdot)(t) : (M, \mathcal{M}) \to [0, \infty]$ is measurable and

$$\int_{T \times M} f(t, m) \mathrm{d}\tau(t, m) = \int_M \int_T f(t, m) \mathrm{d}\beta_i(m)(t) \mathrm{d}\rho(m),$$

with the probability measure τ being given by

$$\tau(H) = \int_M \beta_i(m)(\{t \in T \mid (t, m) \in H\}) \mathrm{d}\rho(t) \quad \forall\, H \in \mathcal{B} \otimes \mathcal{M}.$$

Since for all $T_{\mathcal{B}} \in \mathcal{B},\ M_{\mathcal{M}} \in \mathcal{M}$

$$\begin{aligned}
\pi(T_{\mathcal{B}} \times M_{\mathcal{M}}) &= \int_T \delta_{\sigma_0^1(t)}(\{m \in M \mid (t, m) \in T_{\mathcal{B}} \times M_{\mathcal{M}}\}) \mathrm{d}P(t) \\
&= \int_T \mathbf{1}_{T_{\mathcal{B}}}(t) \delta_{\sigma_0^1(t)}(\{m \in M \mid m \in M_{\mathcal{M}}\}) \mathrm{d}P(t) \\
&= \int_{T_{\mathcal{B}}} \delta_{\sigma_0^1(t)}(M_{\mathcal{M}}) \mathrm{d}P(t) \\
&= (\sigma_0^1 \bullet P)(T_{\mathcal{B}}, M_{\mathcal{M}}) \qquad \text{(by definition)} \\
&= \int_{M_{\mathcal{M}}} \beta_i(m)(T_{\mathcal{B}}) \mathrm{d}\rho(m) \quad \text{(by Bayesian Consistency)} \\
&= \int_M \mathbf{1}_{M_{\mathcal{M}}}(m) \beta_i(m)(\{t \in T \mid t \in T_{\mathcal{B}}\}) \mathrm{d}\rho(m) \\
&= \int_M \beta_i(m)(\{t \in T \mid (t, m) \in T_{\mathcal{B}} \times M_{\mathcal{M}}\}) \mathrm{d}\rho(m) \\
&= \tau(T_{\mathcal{B}} \times M_{\mathcal{M}}),
\end{aligned}$$

π and τ coincide on the set $\{T_{\mathcal{B}} \times M_{\mathcal{M}} \mid T_{\mathcal{B}} \in \mathcal{B},\ M_{\mathcal{M}} \in \mathcal{M}\}$ generating the σ-algebra $\mathcal{B} \otimes \mathcal{M}$. Hence, by [2] Billingsley (1995), Theorem 3.3, $\pi = \tau$[39]. Thus, we have shown for all measurable $f : (T \times M, \mathcal{B} \otimes \mathcal{M}) \to [0, \infty]$ that $\int_M f(\cdot, m) \mathrm{d}\delta_{\sigma_0^1(\cdot)}(m) : (T, \mathcal{B}) \to [0, \infty]$ and $\int_T f(t, \cdot) \mathrm{d}\beta_i(\cdot)(t) : (M, \mathcal{M}) \to [0, \infty]$ are measurable, and

$$\begin{aligned} \int_T f(t, \sigma_0^1(t)) \mathrm{d}P(t) &= \int_T \int_M f(t, m) \mathrm{d}\delta_{\sigma_0^1(t)}(m) \mathrm{d}P(t) \\ &= \int_M \int_T f(s, m) \mathrm{d}\beta_i(m)(s) \mathrm{d}\rho(m) \\ &= \int_T \int_T f(s, \sigma_0^1(t)) \mathrm{d}\beta_i(\sigma_0^1(t))(s) \mathrm{d}P(t), \end{aligned}$$

where the last equality is due to [2] Billingsley (1995), Theorem 16.13. If $f : (T \times M, \mathcal{B} \otimes \mathcal{M}) \to [-U^{max}, U^{max}]$ is measurable, then clearly also $\tilde{f} : (T \times M, \mathcal{B} \otimes \mathcal{M}) \to [0, \infty]$, $(t, m) \mapsto \tilde{f}(t, m) = f(t, m) + U^{max}$ is measurable, and applying the result above to $\tilde{f}$ yields the following claim:

Claim 1 *If $f : (T \times M, \mathcal{B} \otimes \mathcal{M}) \to [-U^{max}, U^{max}]$ is measurable, then*

$$\int_T f(t, \sigma_0^1(t)) \mathrm{d}P(t) = \int_T \psi(\sigma_0^1(t)) \mathrm{d}P(t),$$

where $\psi(m) = \int_T f(s, m) \mathrm{d}\beta_i(m)(s)\ \forall\ m \in M$.

With Claim 1 it is now relatively easy to prove that φ satisfies (IRi).
Since $t \mapsto u_i(d_0, t)$ is measurable, so is $f^1 : (T \times M, \mathcal{B} \otimes \mathcal{M}) \to [-U^{max}, U^{max}]$, $(t, m) \mapsto u_i(d_0, t)$. Applying Claim 1 to f^1 then yields

$$\int_T u_i(d_0, t) \mathrm{d}P(t) = \int_T \psi^1(\sigma_0^1(t)) \mathrm{d}P(t), \tag{4}$$

where $\psi^1(m) = \int_T u_i(d_0, s) \mathrm{d}\beta_i(m)(s)\ \forall\ m \in M$. Next let us apply Claim 1 to g_i. This yields

$$\int_T u_i(\varphi(t), t) \mathrm{d}P(t) = \int_T g_i(t, \sigma_0^1(t)) \mathrm{d}P(t) = \int_T \psi^2(\sigma_0^1(t)) \mathrm{d}P(t), \tag{5}$$

where ψ^2 is given by

$$\begin{aligned} \psi^2(m) &= \int_T g_i(s, m) \mathrm{d}\beta_i(m)(s) \\ &= \int_T u_i(m[\sigma_0^2(s, m, 1^n), \sigma_1^2(m, 1^n), \ldots, \sigma_n^2(m, 1^n)], s) \mathrm{d}\beta_i(m)(s) \end{aligned}$$

[39]Note $\{T_{\mathcal{B}} \times M_{\mathcal{M}} \mid T_{\mathcal{B}} \in \mathcal{B},\ M_{\mathcal{M}} \in \mathcal{M}\}$ is a π-system, i.e. the set is closed under the formation of finite intersections.

for all $m \in M$. Finally, since

$$\psi^2(\sigma_0^1(t)) =$$

$$\int_T u_i(\sigma_0^1(t)[\sigma_0^2(s, \sigma_0^1(t), 1^n), \sigma_1^2(\sigma_0^1(t), 1^n), \ldots, \sigma_n^2(\sigma_0^1(t), 1^n)], s) \mathrm{d}\beta_i(\sigma_0^1(t))(s)$$

$$\geq \int_T u_i(d_0, s) \mathrm{d}\beta_i(\sigma_0^1(t))(s) \;=\; \psi^1(\sigma_0^1(t))$$

for all $t \in T$ (this is imposed by the sequential rationality of (σ, β) for agent i and the assumption that $\sigma_0^1(t)$ is accepted by all agents for all $t \in T$), from equation (4) and equation (5) it follows that

$$\begin{aligned} \int_T u_i(\varphi(t), t) \mathrm{d}P(t) &= \int_T \psi^2(\sigma_0^1(t)) \mathrm{d}P(t) \\ &\geq \int_T \psi^1(\sigma_0^1(t)) \mathrm{d}P(t) \;=\; \int_T u_i(d_0, t) \mathrm{d}P(t). \end{aligned}$$

Now suppose σ_0^1 is an elementary mechanism proposal strategy, i.e. there exists a countable index set N, and for each $k \in N$ some Borel set T_k and a game form $m_k \in M$, such that $\{T_k\}_{k \in N}$ is a partition of T, $m_k \neq m_l \;\forall\; k \neq l$, and such that for all $k \in N$, $t \in T_k$: $\sigma_0^1(t) = m_k$. Since m_l is accepted by all agents for all $l \in N$, we know from the sequential rationality for the agents that for all agents i and for all $l \in N$ with $P(T_l) > 0$

$$\begin{aligned} & \frac{1}{P(T_l)} \int_{T_l} u_i(d_0, t) \mathrm{d}P(t) \\ = & \int_{T_l} u_i(d_0, t) \mathrm{d}\beta_i(m_l)(t) + \int_{T \setminus T_l} u_i(d_0, t) \mathrm{d}\beta_i(m_l)(t) \\ = & \int_T u_i(d_0, t) \mathrm{d}\beta_i(m_l)(t) \\ \leq & \int_T u_i(m_l[\sigma_0^2(t, m_l, 1^n), \sigma_1^2(m_l, 1^n), \ldots, \sigma_n^2(m_l, 1^n)], t) \mathrm{d}\beta_i(m_l)(t) \\ & = \int_{T_l} u_i(m_l[\sigma_0^2(t, m_l, 1^n), \sigma_1^2(m_l, 1^n), \ldots, \sigma_n^2(m_l, 1^n)], t) \mathrm{d}\beta_i(m_l)(t) \\ & \quad + \int_{T \setminus T_l} u_i(m_l[\sigma_0^2(t, m_l, 1^n), \sigma_1^2(m_l, 1^n), \ldots, \sigma_n^2(m_l, 1^n)], t) \mathrm{d}\beta_i(m_l)(t) \\ & = \frac{1}{P(T_l)} \int_{T_l} u_i(m_l[\sigma_0^2(t, m_l, 1^n), \sigma_1^2(m_l, 1^n), \ldots, \sigma_n^2(m_l, 1^n)], t) \mathrm{d}P(t), \end{aligned}$$

where the first and the last equality are due to the fact that $\beta_i(m_l) = \frac{P(\cdot)}{P(T_l)}$ on the set T_l, and $\beta_i(m_l)$ is identical to zero on the set $T \setminus T_l$, as can be seen

from equation (3).
Thus, we can deduce for all $l \in N$

$$\int_{T_l} u_i(d_0, t)\mathrm{d}P(t) \leq \\ \int_{T_l} u_i(m_l[\sigma_0^2(t, m_l, 1^n), \sigma_1^2(m_l, 1^n), \ldots, \sigma_n^2(m_l, 1^n)], t)\mathrm{d}P(t)$$

(obviously, this is also true if $P(T_l) = 0$, because then both sides of the inequality equal zero). Therefore, it follows for all $i = 1, \ldots, n$:

$$\begin{aligned}
& \int_T u_i(\varphi(t), t)\mathrm{d}P(t) \\
= & \int_T \sum_{k \in N} \mathbf{1}_{T_k}(t) u_i(m_k[\sigma_0^2(t, m_k, 1^n), \sigma_1^2(m_k, 1^n), \ldots, \sigma_n^2(m_k, 1^n)], t)\mathrm{d}P(t) \\
= & \sum_{k \in N} \int_T \mathbf{1}_{T_k}(t) u_i(m_k[\sigma_0^2(t, m_k, 1^n), \sigma_1^2(m_k, 1^n), \ldots, \sigma_n^2(m_k, 1^n)], t)\mathrm{d}P(t) \\
= & \sum_{k \in N} \underbrace{\int_{T_k} u_i(m_k[\sigma_0^2(t, m_k, 1^n), \sigma_1^2(m_k, 1^n), \ldots, \sigma_n^2(m_k, 1^n)], t)\mathrm{d}P(t)}_{\geq \int_{T_k} u_i(d_0, t)\mathrm{d}P(t)} \\
\geq & \sum_{k \in N} \int_T \mathbf{1}_{T_k}(t) u_i(d_0, t)\mathrm{d}P(t) \\
& = \int_T \sum_{k \in N} \mathbf{1}_{T_k}(t) u_i(d_0, t)\mathrm{d}P(t) \; = \; \int_T u_i(d_0, t)\mathrm{d}P(t) \; = \; U_i^0.
\end{aligned}$$

So φ is (IRi) for all agents i.
The second and the third last equality above clearly require some justification. First, since (σ, β) satisfies the measurability restriction for a pure-strategy PBE, we know that

$$t \mapsto u_i(m_k[\sigma_0^2(t, m_k, 1^n), \sigma_1^2(m_k, 1^n), \ldots, \sigma_n^2(m_k, 1^n)], t)$$

is measurable for all $i = 1, \ldots, n$, $k \in N$. Hence also

$$t \mapsto \mathbf{1}_{T_k}(t) u_i(m_k[\sigma_0^2(t, m_k, 1^n), \sigma_1^2(m_k, 1^n), \ldots, \sigma_n^2(m_k, 1^n)], t)$$

is measurable for all $i = 1, \ldots, n$, $k \in N$[40], and therefore

$$t \mapsto \sum_{k \in \widetilde{N}} \mathbf{1}_{T_k}(t) u_i(m_k[\sigma_0^2(t, m_k, 1^n), \sigma_1^2(m_k, 1^n), \ldots, \sigma_n^2(m_k, 1^n)], t)$$

[40]Suppose $f : T \to [-U^{max}, U^{max}]$ is measurable, and $S \in \mathcal{B}$. Then, for any Borel set $B \subseteq \mathbb{R}$, $(\mathbf{1}_S(t)f(t))^{-1}(B) = \{t \in T \mid \mathbf{1}_S(t)f(t) \in B\} = \{t \in S \mid f(t) \in B\} \cup \{t \in T \setminus S \mid 0 \in B\}$. Thus, if $0 \in B$, $(\mathbf{1}_S(t)f(t))^{-1}(B) = [f^{-1}(B) \cup (T \setminus S)] \in \mathcal{B}$. On the other hand, if $0 \notin B$, then $(\mathbf{1}_S(t)f(t))^{-1}(B) = f^{-1}(B) \in \mathcal{B}$. Hence, $\mathbf{1}_S(\cdot)f(\cdot)$ is measurable.

is measurable for all $i = 1, \ldots, n$, and all finite subsets $\widetilde{N}$ of N. Finally, from Lebesgue's theorem of dominated convergence (c.f. [2] Billingsley (1995), Theorem 16.4) we can deduce that

$$t \mapsto \sum_{k \in N} \mathbf{1}_{T_k}(t) u_i(m_k[\sigma_0^2(t, m_k, 1^n), \sigma_1^2(m_k, 1^n), \ldots, \sigma_n^2(m_k, 1^n)], t)$$

is measurable for all $i = 1, \ldots, n$ (note, $|\sum_{k \in \widetilde{N}} \mathbf{1}_{T_k}(t) u_i(m_k[\sigma_0^2(t, m_k, 1^n), \sigma_1^2(m_k, 1^n), \ldots, \sigma_n^2(m_k, 1^n)], t)| \leq U^{max} \; \forall \; t \in T$ for all finite subsets $\widetilde{N}$ of N, and $t \mapsto U^{max}$ is clearly integrable with respect to P), and that the second equality above holds. Analogously, it can be shown that

$$\sum_{k \in N} \int_T \mathbf{1}_{T_k}(t) u_i(d_0, t) \mathrm{d}P(t) = \int_T \sum_{k \in N} \mathbf{1}_{T_k}(t) u_i(d_0, t) \mathrm{d}P(t).$$

<u>φ is not strictly dominated at any $t \in T$</u>

Suppose there exists some allocation rule $\psi : T \to D$, $\psi \in M$, and some $\tau \in T$ such that φ is strictly dominated by ψ at τ, i.e. ψ is (IC), $t \mapsto u_i(\psi(t), t)$ is measurable for all i, $u_i(\psi(t), t) > u_i(d_0, t) \; \forall \; t \in T, \; \forall \; i$, and $u_0(\psi(\tau), \tau) > u_0(\varphi(\tau), \tau)$. Since $u_i(\psi(t), t) > u_i(d_0, t) \; \forall \; t \in T$, it follows that $\int_T u_i(\psi(t), t) \mathrm{d}\beta_i(\psi)(t) > \int_T u_i(d_0, t) \mathrm{d}\beta_i(\psi)(t)$, irrespective of how agent i has updated his prior belief to $\beta_i(\psi)$ $(i = 1, \ldots, n)$[41]. Thus, by sequential rationality for agent i, we have $\sigma_i^1(\psi) = 1$ for all agents i[42]. So if the principal proposes ψ, then this allocation rule is accepted by all agents. Therefore, by the sequential rationality of (σ, β) for the principal we can deduce

[41] Though this result is absolutely intuitive, it is not so straight forward to prove this in a direct way. I would like to thank Jens Wannenwetsch for pointing out to me the following result, with which an indirect proof is easily constructed: If $f : (T, \mathcal{B}) \to \mathbb{R}$ is measurable, $f \geq 0$, and $\int_T f \mathrm{d}\mu = 0$ (for some measure μ), then $f = 0$ almost everywhere (see [14] Munroe (1971), Theorem 25.7). Now suppose $\int_T [u_i(\psi(t), t) - u_i(d_0, t)] \mathrm{d}\beta_i(\psi)(t) = 0$ ($\int_T [u_i(\psi(t), t) - u_i(d_0, t)] \mathrm{d}\beta_i(\psi)(t) < 0$ is clearly not possible), then it follows that $u_i(\psi(t), t) - u_i(d_0, t)$ is zero almost everywhere, i.e. $\beta_i(\psi)(T) = 0$ – a contradiction.

[42] This argument is certainly true if $n = 1$. If, however, $n > 1$, then – as already indicated at the introduction of the model – the argument is flawed, since $\sigma_i^1(\psi) = 0$ might be optimal despite the fact that $\int_T u_i(\psi(t), t) \mathrm{d}\beta_i(\psi)(t) > \int_T u_i(d_0, t) \mathrm{d}\beta_i(\psi)(t)$, just because some other agent rejects ψ (even though that agent, as well, would actually prefer ψ compared to the status quo). So *mutual* rejection of ψ may turn out to be optimal due to a coordination failure. The assumption made in footnote No. 17 rules out such coordination failures.

$$\begin{aligned} u_0(\psi(\tau),\tau) \ &\leq \ u_0(\sigma_0^1(\tau)[\sigma_0^2(\tau,\sigma_0^1(\tau),1^n),\sigma_1^2(\sigma_0^1(\tau),1^n),\dots,\sigma_n^2(\sigma_0^1(\tau),1^n)],\tau) \\ &= \ u_0(\varphi(\tau),\tau) \ < \ u_0(\psi(\tau),\tau), \end{aligned}$$

a contradiction.
This proves that φ is not strictly dominated for any $t \in T$.

The first part of this proof has shown that any regular pure-strategy PBE (σ, β) implements some allocation rule φ which is (IC), (IRP), (IRi) for all i, and which is not strictly dominated at any $t \in T$ by any $\psi \in M$. We shall prove now that this allocation rule φ is a (pure-strategy) PBE of the contract proposal game if $\varphi \in M$. To do so, define
$\widetilde{\sigma}_0^1 : T \to M,\ t \mapsto \varphi$,
$\widetilde{\sigma}_0^2 : T \times M \times A \to \bigcup_{S^0 \in \mathcal{S}^0} S^0,\ (t,\varphi,a) \mapsto t,\ (t,m,a) \mapsto \sigma_0^2(t,m,a)\ \forall\ m \neq \varphi$,
$\widetilde{\sigma}_i^1 : M \to \{0,1\},\ \varphi \mapsto 1,\ m \mapsto \sigma_i^2(m)\ \forall\ m \neq \varphi\ (i = 1,\dots,n)$,
$\widetilde{\sigma}_i^2 : M \times A \to \bigcup_{S^i \in \mathcal{S}^i} S^i,\ (\varphi,a) \mapsto s_\varphi^i \in S_\varphi^i,\ (m,a) \mapsto \sigma_i^2(m,a)\ \forall\ m \neq \varphi$ $(i = 1,\dots,n)$, and
$\widetilde{\beta}_i : M \to \mathcal{P}(T),\ \varphi \mapsto P,\ m \mapsto \beta_i(m)\ \forall\ m \neq \varphi\ (i = 1,\dots,n)$.

Measurability
For all agents i, the function

$$t \mapsto u_i(m[\widetilde{\sigma}_0^2(t,m,1^n),\widetilde{\sigma}_1^2(m,1^n),\dots,\widetilde{\sigma}_n^2(m,1^n)],t)$$

is measurable if $m = \varphi$ (since $u_i(\varphi[\widetilde{\sigma}_0^2(t,\varphi,1^n),\widetilde{\sigma}_1^2(\varphi,1^n),\dots,\widetilde{\sigma}_n^2(\varphi,1^n)],t) = u_i(\varphi(t),t)$, and φ is (IRi)). If $m \in M,\ m \neq \varphi$, then this function is given by $t \mapsto u_i(m[\sigma_0^2(t,m,1^n),\sigma_1^2(m,1^n),\dots,\sigma_n^2(m,1^n)],t)$, which is measurable by the measurability-assumption on (σ,β).

Sequential Rationality for the Principal
For the sequential rationality of the principal we first have to verify that

$$u_0(\varphi(t),t) \geq u_0(m[\widetilde{\sigma}_0^2(t,m,1^n),\widetilde{\sigma}_1^2(m,1^n),\dots,\widetilde{\sigma}_n^2(m,1^n)],t)$$

$\forall\ t \in T,\ m \in M$ with $\widetilde{\sigma}_i^1(m) = 1\ \forall\ i = 1,\dots,n$.
If $m = \varphi$, then both sides of the inequality are the same.
So suppose $m \neq \varphi$. The left hand side of the inequality is equal to

$$u_0(\sigma_0^1(t)[\sigma_0^2(t,\sigma_0^1(t),1^n),\sigma_1^2(\sigma_0^1(t),1^n),\dots,\sigma_n^2(\sigma_0^1(t),1^n)],t),$$

and for $m \neq \varphi$ the right hand side reduces to

$$u_0(m[\sigma_0^2(t,m,1^n),\sigma_1^2(m,1^n),\dots,\sigma_n^2(m,1^n)],t),$$

hence, the inequality is implied by the sequential rationality of (σ, β). Moreover, we have to check that $u_0(\varphi(t), t) \geq u_0(d_0, t)\ \forall\ t \in T$. But this is clearly satisfied by (IRP).
Finally, we have to ensure that

$$\begin{aligned} & u_0(m[\widetilde{\sigma}_0^2(t, m, 1^n), \widetilde{\sigma}_1^2(m, 1^n), \ldots, \widetilde{\sigma}_n^2(m, 1^n)], t) \\ \geq \ & u_0(m[s_0, \widetilde{\sigma}_1^2(m, 1^n), \ldots, \widetilde{\sigma}_n^2(m, 1^n)], t)\ \forall\ t \in T,\ m \in M,\ s_0 \in S_m^0. \end{aligned}$$

If $m = \varphi$, then the left hand side is equal to $u_0(\varphi(t), t)$, and the right hand side is equal to $u_0(\varphi(s_0), t)\ \forall\ s_0 \in S_\varphi^0 = T$, so the inequality is implied by the incentive-compatibility of φ. If $m \neq \varphi$, then the inequality is implied again by the sequential rationality of (σ, β).

Sequential Rationality for Agent i

Like with proving that $(\widetilde{\sigma}, \widetilde{\beta})$ is sequential rational for the principal, most of the sequential rationality of $(\widetilde{\sigma}, \widetilde{\beta})$ for the agents can be directly deduced from the corresponding characteristics of (σ, β). Choose some $i \in \{1, \ldots, n\}$ and let us go through the different conditions one at a time.

1. $\int_T u_i(m[\widetilde{\sigma}_0^2(t, m, 1^n), \widetilde{\sigma}_1^2(m, 1^n), \ldots, \widetilde{\sigma}_n^2(m, 1^n)], t) \mathrm{d}\widetilde{\beta}_i(m)(t) \geq$
   $\int_T u_i(d_0, t) \mathrm{d}\widetilde{\beta}_i(m)(t)\ \forall\ m \in M$ satisfying $\widetilde{\sigma}_j^1(m) = 1\ \forall\ j = 1, \ldots, n$.
   If $m \neq \varphi$, this condition follows from (σ, β) being sequential rational for agent i. If $m = \varphi$, then it follows from φ being (IRi).

2. $\int_T u_i(m[\widetilde{\sigma}_0^2(t, m, 1^n), \widetilde{\sigma}_1^2(m, 1^n), \ldots, \widetilde{\sigma}_n^2(m, 1^n)], t) \mathrm{d}\widetilde{\beta}_i(m)(t) \leq$
   $\int_T u_i(d_0, t) \mathrm{d}\widetilde{\beta}_i(m)(t)$
   for all $m \in M$ satisfying $\widetilde{\sigma}_j^1(m) = 1\ \forall\ j \neq i,\ \widetilde{\sigma}_i^1(m) = 0$.
   If $m \neq \varphi$, then again the condition follows from (σ, β) being sequential rational for agent i. If $m = \varphi$, then $\widetilde{\sigma}_i^1(m) = 1$, and thus nothing is to show here.

3. $\int_T u_i(m[\widetilde{\sigma}_0^2(t, m, 1^n), \widetilde{\sigma}_1^2(m, 1^n), \ldots, \widetilde{\sigma}_n^2(m, 1^n)], t) \mathrm{d}\widetilde{\beta}_i(m)(t) \geq$
   $\int_T u_i(m[\widetilde{\sigma}_0^2(t, m, 1^n), \widetilde{\sigma}_1^2(m, 1^n), \ldots, \widetilde{\sigma}_{i-1}^2(m, 1^n), s_i,$
   $\widetilde{\sigma}_{i+1}^2(m, 1^n), \ldots, \widetilde{\sigma}_n^2(m, 1^n)], t) \mathrm{d}\widetilde{\beta}_i(m)(t)$
   for all $m \in M$ and for all $s_i \in S_m^i$ which lead to
   $t \mapsto u_i(m[\widetilde{\sigma}_0^2(t, m, 1^n), \widetilde{\sigma}_1^2(m, 1^n), \ldots, \widetilde{\sigma}_{i-1}^2(m, 1^n), s_i,$
   $\widetilde{\sigma}_{i+1}^2(m, 1^n), \ldots, \widetilde{\sigma}_n^2(m, 1^n)], t)$
   being measurable.

Once again, for $m \neq \varphi$ this requirement is fulfilled, since (σ, β) is sequential rational for agent i. If $m = \varphi$, then there is nothing to show, as for $m = \varphi$, agent i does not have any choice of action in the mechanism execution stage.

Bayesian Consistency

Clearly, $(\widetilde{\sigma}, \widetilde{\beta})$ is Bayesian consistent, and both equation (2) and equation (3) hold: Equation (3) holds, as $\{T\}$ is the appropriate partition of T to be used for equation (3), and $\beta_i(\varphi) = P$. To verify that

$$P(T_{\mathcal{B}} \cap \widetilde{\sigma}_0^{-1}(M_{\mathcal{M}})) = \int_{\widetilde{\sigma}_0^{-1}(M_{\mathcal{M}})} \widetilde{\beta}_i(\widetilde{\sigma}_0^1(t))(T_{\mathcal{B}}) \mathrm{d}P(t)$$

holds for all agents i and for all $T_{\mathcal{B}} \in \mathcal{B}$, $M_{\mathcal{M}} \in \mathcal{M}$, note that the right hand side of the equation above reduces to

$$\int_{\widetilde{\sigma}_0^{-1}(M_{\mathcal{M}})} P(T_{\mathcal{B}}) \mathrm{d}P(t) = P(T_{\mathcal{B}}) P(\widetilde{\sigma}_0^{-1}(M_{\mathcal{M}})) = \begin{cases} P(T_{\mathcal{B}}) & : \quad \varphi \in M_{\mathcal{M}} \\ 0 & : \quad \varphi \notin M_{\mathcal{M}} \end{cases}.$$

Thus Bayesian consistency of $(\widetilde{\sigma}, \widetilde{\beta})$ follows from the fact that $T_{\mathcal{B}} \cap \widetilde{\sigma}_0^{-1}(M_{\mathcal{M}}) = T_{\mathcal{B}}$ if $\varphi \in M_{\mathcal{M}}$, and $T_{\mathcal{B}} \cap \widetilde{\sigma}_0^{-1}(M_{\mathcal{M}}) = \emptyset$, otherwise.

Hence, $(\widetilde{\sigma}, \widetilde{\beta})$ is a PBE. □

Proof of Theorem 2

Consider some allocation rule $\varphi : T \to D$ in M satisfying (IC), (IRP), and (IRi) for all $i \in \{1, \ldots, n\}$. We first show that φ can be implemented through a Bayesian equilibrium of the contract proposal game, i.e. we show that there exist some strategies σ and some updated beliefs β such that (σ, β) is sequentially rational on the equilibrium path and Bayesian consistent, and such that (σ, β) implements φ.

Define $\sigma_0^1 : T \to M$ by $\sigma_0^1(t) = \varphi$ for all $t \in T$, and let $\sigma_0^2(t, m, a)$ be such that $\sigma_0^2(t, \varphi, a) = t$ for all $t \in T,\ a \in A$. Furthermore, for $i \in \{1, \ldots, n\}$ define $\sigma_i^1 : M \to \{0, 1\}$ by $\sigma_i^1(\varphi) = 1$ and $\sigma_i^1(m) = 0$ for all $m \in M$, $m \neq \varphi$, and let σ_i^2 be any admissible strategy for agent i for the mechanism execution stage of the contract proposal game. Finally, let $\beta_i : M \to \mathcal{P}(T)$ be any updated belief for agent i with $\beta_i(\varphi) = P$. Clearly, σ_0^1 is optimal for the principal, since σ_0^1 yields her $u_0(\varphi(\sigma_0^2(t, \varphi, 1^n)), t) = u_0(\varphi(t), t) \geq u_0(d_0, t)\ \forall\ t \in T$ (by (IRP)), whereas the announcement of any other $m \in M$ leads to a rejection by the agents and thus would yield only $u_0(d_0, t)$. Moreover, σ_0^2 is optimal on the equilibrium path, since φ is incentive-compatible. σ_i^1 is optimal on the equilibrium path for all agents i, since accepting φ yields agent i an expected utility of $\int_T u_i(\varphi(\sigma_0^2(t,\varphi,1^n)),t)\mathrm{d}\beta_i(\varphi)(t) = \int_T u_i(\varphi(t),t)\mathrm{d}P(t) \geq U_i^0$ (by (IRi)), and σ_i^2 is optimal on the equilibrium path since the agents do not have a choice to make in the mechanism execution stage if φ is announced by the principal. Since β is obviously Bayesian consistent, it follows that (σ,β) is a Bayesian equilibrium which implements φ.

Now suppose that

$$M \subseteq \{m : T \to D \mid \forall_{t\in T}\ \exists_{\tau_t^m \in T}\ \forall_{s\in T}\ u_0(m(\tau_t^m),t) \geq u_0(m(s),t) \text{ and } t \mapsto u_i(m(\tau_t^m),t) \text{ is measurable } \forall\ i>0\}$$

and that $\varphi \in M$ is incentive-compatible, individual rational, and not strictly dominated at any $t \in T$ by any $\psi : T \to D$. Consider the following strategies and updated beliefs: $\sigma_0^1 : T \to M,\ t \mapsto \varphi$; $\sigma_0^2 : T \times M \times A \to T$, $(t,\varphi,a) \mapsto t$, and $(t,m,a) \mapsto \tau_t^m\ \forall\ m \neq \varphi$, where for all $m \in M$ $\tau_\cdot^m : T \to T$ is such that $u_0(m(\tau_t^m),t) \geq u_0(m(s),t)\ \forall\ s \in T$ and $t \mapsto u_i(m(\tau_t^m),t)$ is measurable for all $i > 0$ (such $\tau_\cdot^m$ exist for all $m \in M$ by the restriction made on M). Furthermore, define $L := \{m \in M \mid u_0(m(\tau_t^m),t) \leq u_0(\varphi(t),t)\ \forall\ t \in T\}$. For all $m \in M \setminus L$ there exists some $i_m \in \{1,\ldots,m\}$, $t_m \in T$ with $u_{i_m}(m(\tau_{t_m}^m),t_m) \leq u_{i_m}(d_0,t_m)$, otherwise φ would be strictly dominated by $t \mapsto m(\tau_t^m)$ at some $t \in T$. For all $i = 1,\ldots,n$, set $\beta_i(m) = P$ if $m \in L \cup \{\varphi\}$, $\beta_i(m) = \delta_{t_m}$ if $m \notin L$, $m \neq \varphi$. Moreover, for all $m \in M$ and all i set $\sigma_i^1(m) = 1$ if $\int_T u_i(m(\sigma_0^2(t,m,1^n)),t)\mathrm{d}\beta_i(m)(t) > \int_T u_i(d_0,t)\mathrm{d}\beta_i(m)(t)$ or $m = \varphi$, and set $\sigma_i^1(m) = 0$ otherwise. As M is restricted to allocation rules, the agents do not have any choice of action in the mechanism execution stage, so we do not have to specify σ_i^2 for any agent i.

We shall prove that (σ,β) as defined above satisfies the four conditions for a PBE specified in Subsection 2.3. As $\sigma_0^1(t) = \varphi$ for all $t \in T$ and $\sigma_i^1(\varphi) = 1$ for all $i \in \{1,\ldots,n\}$, this demonstrates that φ is a PBE of the contract proposal game.

Measurability

σ_0^2 is constructed such that $t \mapsto u_i(m(\sigma_0^2(t,m,1^n)),t) = u_i(m(\tau_t^m),t)$ is measurable for all $i \in \{1,\ldots,n\}$, for all $m \in M$, $m \neq \varphi$. $t \mapsto u_i(\varphi(\sigma_0^2(t,\varphi,1^n)),t)$ $= u_i(\varphi(t),t)$ is measurable for all i due to (IRi).

Sequential Rationality for the Principal

$$\begin{aligned} u_0(\sigma_0^1(t)(\sigma_0^2(t,\sigma_0^1(t),1^n)),t) &= u_0(\varphi(t),t) \\ &\geq u_0(m(\tau_t^m),t) = u_0(m(\sigma_0^2(t,m,1^n)),t) \end{aligned}$$

for all $m \in L$. If $m \in M \setminus L$ then $\int_T u_{i_m}(m(\sigma_0^2(t,m,1^n)),t)\mathrm{d}\beta_{i_m}(m)(t) = \int_T u_{i_m}(m(\tau_t^m),t)\mathrm{d}\delta_{t_m}(t) = u_{i_m}(m(\tau_{t_m}^m),t_m) \leq u_{i_m}(d_0,t_m) = \int_T u_{i_m}(d_0,t)\,\mathrm{d}\delta_{t_m}(t) = \int_T u_{i_m}(d_0,t)\mathrm{d}\beta_{i_m}(m)(t)$, thus $\sigma_{i_m}^1(m) = 0$ (note, $m \neq \varphi$). Since φ is (IRP), it follows that σ_0^1 is sequentially rational. σ_0^2 is sequentially rational by the definition of $\tau_\cdot^m$ and by the incentive-compatibility of φ.

Sequential Rationality for Agent i

$\sigma_i^1(m) = 1$ if and only if the expected utility agent i derives from m is greater than the expected utility i derives from the status quo or if $m = \varphi$. Since φ satisfies (IRi), this yields the sequential rationality of σ_i^1.

Bayesian Consistency

(σ, β) is Bayesian consistent since the principal does not reveal any information by announcing φ, and $\beta_i(\varphi) = P$ for all i. □

Proof of Lemma 1

Suppose (σ, β) is a pure-strategy perfect Bayesian pooling equilibria of the mechanism signaling game involving play of 'accept' by all agents. The mechanism signaling game essentially corresponds to the contract proposal game with $M = M^*$, and a pooling PBE of the mechanism signaling game is clearly an equilibrium with an elementary mechanism proposal strategy. Hence, by Theorem 1, we know that (σ, β) implements an allocation rule which is (IC), (IRP), (IRi) for all i, and which is not strictly dominated (a direct proof which does not rely on Theorem 1 is easily constructed, as well).

Now suppose $\varphi : T \to D$ is (IC), (IRP), (IRi) for all i, and not strictly dominated at any $t \in T$. From (IC) and (IRi) we know that $\varphi \in M^*$. Define $\sigma_0 : T \to M^*$ by $\sigma_0(t) = \varphi \ \forall\ t \in T$. Furthermore, define the set $L := \{\psi \in M^* \mid u_0(\psi(t),t) \leq u_0(\varphi(t),t) \ \forall\ t \in T\}$. For any $\psi \in M^* \setminus L$, there exists some $i_\psi \in \{1,\ldots,n\}$ and some $t_\psi \in T$ such that $u_{i_\psi}(\psi(t_\psi),t_\psi) \leq u_{i_\psi}(d_0,t_\psi)$, otherwise φ would be strictly dominated by ψ at some $t \in T$. Now define for all agents $\beta_i(\psi) = P$ if $\psi \in L$ and $\beta_i(\psi) = \delta_{t_\psi}$ if $\psi \in M^* \setminus L$. Finally, define agent i's strategy as follows: $\sigma_i(\psi) = 1$ if $\int_T u_i(\psi(t),t)\mathrm{d}\beta_i(\psi)(t) > \int_T u_i(d_0,t)\mathrm{d}\beta_i(\psi)(t)$ or if $\psi = \varphi$, and $\sigma_i(\psi) = 0$ otherwise ($\psi \in M^*$).

Clearly, σ represents a set of pure strategies for the mechanism signaling game such that the allocation rule $\sigma_0(t) = \varphi$ is accepted by all agents for any $t \in T$. We shall show now that (σ, β) is a PBE (hence it is a pooling PBE). (σ, β) is Bayesian consistent, since $\beta_i(\varphi) = P \ \forall\ i = 1, \ldots, n$. Moreover, (σ, β) is sequential rational for all agents, because any allocation rule proposed by the principal is accepted if and only if it yields the agent an expected utility greater than the expected utility derived from the status quo or if φ is proposed (which yields an expected utility greater than or equal to the expected utility derived from the status quo, given the updated belief $\beta_i(\varphi) = P$, since φ is (IRi) for all i). To verify that (σ, β) is sequential rational for the principal, consider some allocation rule $\psi \neq \varphi$. If $\psi \in L$, then the principal does not have any reason to propose ψ rather than φ. If $\psi \notin L$, then the expected utility agent i_ψ derives from ψ is less than or equal to $\int_T u_{i_\psi}(d_0, t) \mathrm{d}\beta_{i_\psi}(\psi)(t)$ (since his belief is given by $\beta_{i_\psi}(\psi) = \delta_{t_\psi}$, so $\int_T u_{i_\psi}(\psi(t), t) \mathrm{d}\beta_{i_\psi}(\psi)(t) = u_{i_\psi}(\psi(t_\psi), t_\psi) \leq u_{i_\psi}(d_0, t_\psi) = \int_T u_{i_\psi}(d_0, t) \mathrm{d}\beta_{i_\psi}(\psi)(t)$), hence $\sigma_{i_\psi}(\psi) = 0$. Thus it is optimal for the principal not to propose ψ instead of φ, since φ yields her some utility $u_0(\varphi(t), t) \geq u_0(d_0, t)$ (as φ is (IRP)). Hence (σ, β) is a PBE which implements φ[43]. □

Proof of Theorem 3

Let us start assuming that φ is some allocation rule which is supported by a pure-strategy perfect Bayesian pooling equilibrium of the mechanism signaling game with perfect beliefs off the equilibrium path, and which involves 'acceptance' by all agents. Then, by Lemma 1, φ is (IC), (IRP), and (IRi) for all i. So φ satisfies condition (1) for being a perfect solution. Furthermore, consider some incentive-compatible allocation rule ψ satisfying that $t \mapsto u_i(\psi(t), t)$ is measurable for all i, with $S := \{t \in T \mid u_0(\psi(t), t) > u_0(\varphi(t), t)\} \neq \emptyset$. If condition (2) of Definition 7 was not satisfied, then ψ would be accepted by all agents if it was announced (since beliefs off the equilibrium path are perfect)[44]; hence all principals with types in S would strictly benefit from announcing ψ rather than φ, contradicting that φ is an equilibrium allocation rule.

Now let φ be some perfect solution. From condition (1) of Definition 7, $\varphi \in M^*$. Define $\sigma_0 : T \to M^*$ by $\sigma_0(t) = \varphi \ \forall\ t \in T$. Furthermore, define for all agents $\beta_i(\varphi) = P$, and for all allocation rules ψ off the equilibrium

[43]Note that $t \mapsto u_i(m(t), t)$ is measurable for all $i = 1, \ldots, n,\ m \in M^*$, so (σ, β) also satisfies the measurability condition listed in Subsection 2.3.

[44]Here we require that an allocation rule ψ is accepted by agent i if $\int_T u_i(\psi(t), t) \mathrm{d}\beta_i(\psi)(t) > \int_T u_i(d_0, t) \mathrm{d}\beta_i(\psi)(t)$.

path, let $\beta_i(\psi)$ be perfect, i.e. define $S := \{t \in T \mid u_0(\psi(t), t) > u_0(\varphi(t), t)\}$ and choose some set $\widehat{S} \in \text{argmax}\{P(\widetilde{S}) \mid S \supseteq \widetilde{S} \in \mathcal{B}\}$; if $P(\widehat{S}) > 0$, then set $\beta_i(\psi)(T_{\mathcal{B}}) = \frac{P(T_{\mathcal{B}} \cap \widehat{S})}{P(\widehat{S})}$ $\forall$ $T_{\mathcal{B}} \in \mathcal{B}$; otherwise, by condition (2) of Definition 7, there exists some $i_\psi \in \{1, \ldots, n\}$ and some $t_\psi \in S$ such that $u_{i_\psi}(\psi(t_\psi), t_\psi) \leq u_{i_\psi}(d_0, t_\psi)$; define $\beta_i(\psi) = \delta_{t_\psi}$ (provided $S \neq \emptyset$). If $S = \emptyset$, then define $\beta_i(\psi) = P$. Finally, let agent i's strategy be defined as follows: $\sigma_i(\psi) = 1$ if $\int_T u_i(\psi(t), t)\mathrm{d}\beta_i(\psi)(t) > \int_T u_i(d_0, t)\mathrm{d}\beta_i(\psi)(t)$ or if $\psi = \varphi$, and $\sigma_i(\psi) = 0$, otherwise (for $\psi \in M^*$).

Clearly, (σ, β) represents a set of pure strategies for the mechanism signaling game such that the allocation rule $\sigma_0(t) = \varphi$ is accepted by all agents for any $t \in T$. We shall show that (σ, β) is a PBE of the mechanism signaling game which, of course, implements the allocation rule φ (thus, (σ, β) is a (pure-strategy) PBE with perfect beliefs off the equilibrium path).

(σ, β) is Bayesian consistent, since $\beta_i(\varphi) = P \; \forall \, i = 1, \ldots, n$. Moreover, (σ, β) is sequential rational for all agents, because any allocation rule proposed by the principal is accepted if and only if it yields the agent an expected utility greater than his expected utility from the status quo or if $\psi = \varphi$ (note, any perfect solution is (IRi) for all i, so φ yields the agents an expected utility greater than or equal to their expected utilities from the status quo). To verify that (σ, β) is sequential rational for the principal, consider some allocation rule $\psi \neq \varphi$. If $S = \{t \in T \mid u_0(\psi(t), t) > u_0(\varphi(t), t)\} = \emptyset$, then the principal does not have any incentive to deviate from φ to ψ (irrespective of the principal's type). So suppose $S \neq \emptyset$. If $P(\widehat{S}) > 0$, then by condition (2) of Definition 7, there exists some $i \in \{1, \ldots, n\}$ with $\int_T u_i(\psi(t), t)\mathrm{d}\beta_i(\psi)(t) = \frac{1}{P(\widehat{S})} \int_{\widehat{S}} u_i(\psi(t), t)\mathrm{d}P(t) \leq \frac{1}{P(\widehat{S})} \int_{\widehat{S}} u_i(d_0, t)\mathrm{d}P(t) = \int_T u_i(d_0, t)\mathrm{d}\beta_i(\psi)(t)$. Hence, $\sigma_i(\psi) = 0$. If $P(\widehat{S}) = 0$, then the expected utility agent i_ψ derives from ψ is less than or equal to $\int_T u_{i_\psi}(d_0, t)\mathrm{d}\beta_{i_\psi}(\psi)(t)$ (since his belief is given by $\beta_{i_\psi}(\psi) = \delta_{t_\psi}$, so $\int_T u_{i_\psi}(\psi(t), t)\mathrm{d}\beta_{i_\psi}(\psi)(t) = u_{i_\psi}(\psi(t_\psi), t_\psi) \leq u_{i_\psi}(d_0, t_\psi) = \int_T u_{i_\psi}(\psi(t), t)\mathrm{d}\beta_{i_\psi}(\psi)(t))$, hence, again, $\sigma_{i_\psi}(\psi) = 0$. Thus, the principal does not benefit from proposing ψ rather than φ (since φ yields her a utility of $u_0(\varphi(t), t) \geq u_0(d_0, t)$ due to φ being (IRP)). Hence, (σ, β) is a PBE which implements φ. □

Proof of Theorem 4

Consider some regular pure-strategy PBE (σ, β) of the contract proposal game with perfect beliefs off the equilibrium path. From Theorem 1 we know that (σ, β) implements an allocation rule φ which is (IC), (IRP), and (IRi) for all i. It remains to prove that φ satisfies condition (2) of Definition 7. Suppose, to the contrary, that there exists some incentive-compatible allocation rule $\psi : T \to D$, $\psi \in M$, satisfying that $t \mapsto u_i(\psi(t), t)$ is measurable for all agents i, such that $S = \{t \in T \mid u_0(\psi(t), t) > u_0(\varphi(t), t)\} \neq \emptyset$ and such that condition (2) is violated. Let $\widehat{S} \in \text{argmax}\{P(\widetilde{S}) \mid S \supseteq \widetilde{S} \in \mathcal{B}\}$. If $P(\widehat{S}) > 0$, then ψ violating condition (2) says that for all agents i, $\int_{\widehat{S}} u_i(\psi(t), t)\mathrm{d}P(t) > \int_{\widehat{S}} u_i(d_0, t)\mathrm{d}P(t)$. But since agents use perfect beliefs off the equilibrium path, from $P(\widehat{S}) > 0$ it follows for all $i = 1, \ldots, n$, that $\beta_i(\psi) = \frac{P(\cdot \cap \widehat{S})}{P(\widehat{S})}$, and thus

$$\begin{aligned}\int_T u_i(\psi(t), t)\mathrm{d}\beta_i(\psi)(t) &= \frac{1}{P(\widehat{S})}\int_{\widehat{S}} u_i(\psi(t), t)\mathrm{d}P(t) \\ &> \frac{1}{P(\widehat{S})}\int_{\widehat{S}} u_i(d_0, t)\mathrm{d}P(t) = \int_T u_i(d_0, t)\mathrm{d}\beta_i(\psi)(t).\end{aligned}$$

Therefore, ψ would be accepted by all agents[45], implying that all types of principals in the set S should deviate from equilibrium play to ψ, and thus that (σ, β) is not a PBE.

Similarly, if $P(\widehat{S}) = 0$, then ψ violating condition (2) says that for all agents i and all $t \in S$, $u_i(\psi(t), t) > u_i(d_0, t)$. Thus, all agents would accept ψ and, hence, (σ, β) would not be sequential rational for the principal, so (σ, β) would not be a PBE. □

[45] Here, again, we have to assume that agents accept an allocation rule if it yields them an expected utility greater than their expected utility from the status quo – c.f. footnote No. 42 and footnote No. 44.

Proof of Theorem 5

The proof of Theorem 5 (as well as the theorem itself) is inspired by (the proof of) Theorem 1 in [16] Myerson (1983).

Let φ be some strong solution. Then φ is (IC) and (IRi) for all i, since it is feasible, and φ is (IRP) since it is safe. Now consider some incentive-compatible allocation rule ψ such that $t \mapsto u_i(\psi(t), t)$ is measurable for all i, for which the set $S = \{t \in T \mid u_0(\psi(t), t) > u_0(\varphi(t), t)\}$ is not empty. Note $S \in \mathcal{B}$, since $t \mapsto u_i(\psi(t), t)$ and $t \mapsto u_i(\varphi(t), t)$ are measurable for all i and hence $t \mapsto u_0(\psi(t), t)$ and $t \mapsto u_0(\varphi(t), t)$ are measurable.

Define

$$\varphi^*(t) = \begin{cases} \psi(t) & : \quad t \in S \\ \varphi(t) & : \quad t \in T \setminus S \end{cases}$$

$\Rightarrow t \mapsto u_0(\varphi^*(t), t)$ is measurable as the maximum of two measurable functions (since $t \mapsto u_0(\varphi(t), t)$ and $t \mapsto u_0(\psi(t), t)$ are measurable)

$\Rightarrow t \mapsto u_i(\varphi^*(t), t)$ is measurable for all i.

Furthermore, φ^* is incentive-compatible, since

- $\forall\, t, \tilde{t} \in T \setminus S :$
  $u_0(\varphi^*(t), t) = u_0(\varphi(t), t) \geq u_0(\varphi(\tilde{t}), t) = u_0(\varphi^*(\tilde{t}), t)$, since φ is (IC),

- $\forall\, t, \tilde{t} \in S :$
  $u_0(\varphi^*(t), t) = u_0(\psi(t), t) \geq u_0(\psi(\tilde{t}), t) = u_0(\varphi^*(\tilde{t}), t)$, since ψ is (IC),

- $\forall\, t \in T \setminus S,\ \tilde{t} \in S :$
  $u_0(\varphi^*(t), t) = u_0(\varphi(t), t) \geq u_0(\psi(t), t) \geq u_0(\psi(\tilde{t}), t) = u_0(\varphi^*(\tilde{t}), t)$, since $t \notin S$ and ψ is (IC), and

- $\forall\, t \in S, \tilde{t} \in T \setminus S :$
  $u_0(\varphi^*(t), t) = u_0(\psi(t), t) > u_0(\varphi(t), t) \geq u_0(\varphi(\tilde{t}), t) = u_0(\varphi^*(\tilde{t}), t)$, since $t \in S$ and φ is (IC),

i.e. $\forall\, t, \tilde{t} \in T :\ u_0(\varphi^*(t), t) \geq u_0(\varphi^*(\tilde{t}), t)$.

Moreover, $u_0(\varphi^*(t), t) \geq u_0(\varphi(t), t)\ \forall\, t \in T$ and $u_0(\varphi^*(t), t) > u_0(\varphi(t), t)$ $\forall\, t \in S$. Hence $\varphi^* \notin \mathcal{F}$, because otherwise φ was dominated. But since φ^* is (IC) and since $t \mapsto u_i(\varphi^*(t), t)$ is measurable for all i, it follows that there exists some $j \in \{1, \ldots, n\}$ such that

$$\int_T u_j(\varphi^*(t), t) \mathrm{d}P(t) < U_j^0 = \int_T u_j(d_0, t) \mathrm{d}P(t).$$

Since $S \in \mathcal{B}$, it follows that

$$\begin{aligned}\int_T u_j(d_0,t)\mathrm{d}P(t) \; > \; & \int_T u_j(\varphi^*(t),t)\mathrm{d}P(t) \\ = \; & \int_S u_j(\psi(t),t)\mathrm{d}P(t) \; + \; \int_{T\setminus S} \underbrace{u_j(\varphi(t),t)}_{\geq\, u_j(d_0,t)} \mathrm{d}P(t),\end{aligned}$$

where the greater-or-equal-inequality is due to the fact that φ is safe. Hence

$$\int_S u_j(d_0,t)\mathrm{d}P(t) > \int_S u_j(\psi(t),t)\mathrm{d}P(t).$$

Note if $P(S) = 0$, then both integrals would be equal to zero – a contradiction, so $P(S) > 0$. Now choose some arbitrary $\widehat{S} \in \operatorname{argmax}\{P(\widetilde{S}) \mid S \supseteq \widetilde{S} \in \mathcal{B}\}$. Clearly, $P(\widehat{S}) = P(S) > 0$, and

$$\begin{aligned}\int_{\widehat{S}} u_j(\psi(t),t)\mathrm{d}P(t) \; = \; & \int_S u_j(\psi(t),t)\mathrm{d}P(t) \\ < \; & \int_S u_j(d_0,t)\mathrm{d}P(t) = \int_{\widehat{S}} u_j(d_0,t)\mathrm{d}P(t)\end{aligned}$$

(since $\widehat{S}$ and S differ only by some set of measure zero). Therefore, φ also satisfies condition (2) of Definition 7 for being a perfect solution. As an undominated allocation rule, φ is not dominated by any other feasible allocation rule, in particular it is not dominated by any perfect solution (note, perfect solutions are feasible). Thus, φ is an undominated perfect solution. Next let $\widetilde{\varphi}$ be some other strong solution and consider the set S above for $\psi = \widetilde{\varphi}$. Suppose $S \neq \emptyset$. As above, it follows that the *combined* allocation rule φ^* is (IC) and that $t \mapsto u_i(\varphi^*(t),t)$ is measurable for all i, and therefore that there exists some $j \in \{1,\ldots,n\}$ satisfying $\int_T u_j(\varphi^*(t),t)\mathrm{d}P(t) < \int_T u_j(d_0,t)\mathrm{d}P(t)$. But this is a contradiction, because $u_j(\varphi^*(t),t) \geq u_j(d_0,t)$ no matter if $t \in S$ or $t \in T \setminus S$, since both φ and $\widetilde{\varphi}$ are safe. Hence $S = \emptyset$, i.e. $u_0(\widetilde{\varphi}(t),t) \leq u_0(\varphi(t),t) \ \forall\ t \in T$. Interchanging the roles of φ and $\widetilde{\varphi}$ yields $u_0(\varphi(t),t) \leq u_0(\widetilde{\varphi}(t),t) \ \forall\ t \in T$. □

References

[1] Baron, D.P., and Myerson, R.B. (1982), 'Regulating A Monopolist With Unknown Costs', *Econometrica,* **50**(4), 911–930

[2] Billingsley, P. (1995), 'Probability and Measure', 3rd ed., *John Wiley and Sons, New York*

[3] Beaudry, P. (1994), 'Why an Informed Principal May Leave Rents to an Agent', *International Economic Review,* **35**(4), 821–832

[4] Cho, I.-K., and Kreps, D.M. (1987), 'Signaling Games and Stable Equilibria', *Quarterly Journal of Economics,* **102**(2), 180–221

[5] Da Prato, G., and Zabczyk, J. (1992), 'Stochastic Equations in Infinite Dimensions', Encyclopedia of Mathematics and its Applications, *Cambridge University Press*

[6] Ellsberg, D. (1961), 'Risk, Ambiguity, and the Savage Axioms', *Quarterly Journal of Economics,* **75**, 643–669

[7] Fudenberg, D., and Tirole, J. (1991) 'Perfect Bayesian Equilibrium and Sequential Equilibrium', *Journal of Economic Theory,* **53**, 236–260

[8] Grossman, S.J., and Perry, M. (1986) 'Perfect Sequential Equilibrium', *Journal of Economic Theory,* **39**, 97–119

[9] Kreps, D.M., and Wilson, R. (1982), 'Sequential Equilibria', *Econometrica,* **50**(4), 863–894

[10] Mailath, G.J., Okuno-Fujiwara, M., and Postlewaite, A. (1993) 'Belief-Based Refinements in Signalling Games', *Journal of Economic Theory,* **60**, 241–276

[11] Manelli, A.M. (1996) 'Cheap Talk Sequential Equilibria in Signaling Games', *Econometrica* **64**(4), 917–942

[12] Maskin, E., and Tirole, J. (1990), 'The Principal-Agent Relationship with an Informed Principal: The Case of Private Values', *Econometrica,* **58**(2), 379–409

[13] Maskin, E., and Tirole, J. (1992), 'The Principal-Agent Relationship with an Informed Principal, II: Common Values', *Econometrica,* **60**(1), 1–42

[14] Munroe, M.E. (1971), 'Measure and Integration' *Addison-Wesley, Reading*

[15] Myerson, R.B. (1981), 'Optimal Auction Design', *Mathematics of Operations Research,* **6**(1), 58–73

[16] Myerson, R.B. (1983), 'Mechanism Design by an Informed Principal', *Econometrica,* **51**(6), 1767–1797

[17] Ottaviani, M., and Prat, A. (2001), 'The Value of Public Information in Monopoly', *Econometrica,* **69**(6), 1673–1683

[18] Perea y Monsuwe, A., Jansen, M., and Peters, H. (1997), 'Consistency of assessments in infinite signaling games', *Journal of Mathematical Economics,* **27**, 425–449

[19] Stoughton, N.M., and Talmor, E. (1994), 'A mechanism design approach to transfer pricing by the multinational firm', *European Economic Review,* **38**, 143–170

[20] Stoughton, N.M., and Talmor, E. (1999), 'Managerial Bargaining Power in the Determination of Compensation Contracts and Corporate Investment', *International Economic Review,* **40**(1), 69–93

[21] Tan, G. (1996), 'Optimal Procurement Mechanisms for an Informed Buyer', *Canadian Journal of Economics,* **29**(3), 699–716

[22] Tisljar, R. (2002), 'Selling-Mechanisms for a Model of Bilateral Trade with Adverse Selection', *Working Paper,* Bonn Graduate School of Economics, September 2002

[23] Yilankaya, O. (1999), 'A Note on the Seller's Optimal Mechanism in Bilateral Trade with Two-Sided Incomplete Information', *Journal of Economic Theory,* **87**, 267–271