Review of Economic Studies (1999) **66**, 83–114 0034-6527/99/00050083$02.00

Unforeseen Contingencies and Incomplete Contracts

ERIC MASKIN
Harvard University

and

JEAN TIROLE
IDEI, GREMAQ, CERAS and MIT

First version received April 1997; *Final version accepted October* 1998 (*Eds.*)

We scrutinize the conceptual framework commonly used in the incomplete contract literature. This literature usually assumes that contractual incompleteness is due to the transaction costs of describing—or of even foreseeing—the possible states of nature in advance. We argue, however, that such transaction costs need not interfere with optimal contracting (*i.e.* transaction costs need not be *relevant*), provided that agents can probabilistically forecast their possible future *payoffs* (even if other aspects of the state of the nature cannot be forecast). In other words, all that is required for optimality is that agents be able to perform dynamic programming, an assumption always invoked by the incomplete contract literature. The foregoing optimality result holds very generally provided that parties can commit themselves not to renegotiate. Moreover, we point out that renegotiation may be hard to reconcile with a framework that otherwise presumes perfect rationality. However, even if renegotiation is allowed, the result still remains valid provided that parties are risk averse.

1. INTRODUCTION

Most real contracts are vague or silent on a number of significant matters. For example, in government, the mandates of agencies, courts, or legislators make almost no mention of the many contingencies that may affect the set of feasible options, nor of how the choice among options should reflect those contingencies. Moreover, this "contractual incompleteness" appears to be important to understanding why the structure of government matters. Similarly, incompleteness is often thought to be crucial to explaining the boundaries of the firm, corporate control, and the patent system.

While there has been a flurry of contributions to the economics of property and control rights since the early work of Grossman and Hart (1986) and Hart and Moore (1990), there is no well-accepted incomplete contracting paradigm, unlike for the phenomena of moral hazard and adverse selection. A few papers study very structured environments in a complete-contracting framework to show that in specific instances in which there are many constraints the optimal complete contract reduces to something that looks quite "incomplete": *e.g.* no contract at all (Che–Hausch (1998), Hart–Moore (1999), Segal (1995, 1999)), an authority relationship (Aghion–Tirole (1997)), an option to buy (Nöldeke–Schmidt (1998)), a property right (Maskin–Tirole (1999)), or some other simple decision process (see Tirole (1999) for a discussion). But the bulk of the papers on incomplete contracting (for epistemological reasons or else for simplicity) do not attempt to derive complete contract foundations for the restricted class of contracts they study. Rather, following Williamson (1975, 1985), they invoke *transaction* costs (difficulty of foreseeing contingencies, cost of writing numerous contingencies, cost of enforcement by

a court) to motivate the restriction to a simple class of contracts such as property rights or authority.

This paper is a methodological discussion of the incomplete contract literature. We question existing modelling by pointing out a tension between two important features of the literature: the postulation of significant transaction costs and the use of dynamic programming. Roughly, we argue that the rationality needed to perform dynamic programming is in standard models strong enough to ensure that transaction costs are *irrelevant*. More specifically, we show that even if such costs prevent agents from describing physical contingencies *ex ante*, and in the class of models on which the incomplete contracting literature usually focuses, they do not constrain the set of payoffs that can be reached through contracting in the absence of contract renegotiation. Thus, although we certainly acknowledge that transaction costs matter in reality, we believe that more attention needs to be devoted to the conceptual underpinnings of incomplete contract models.

The basic idea behind the *irrelevance theorem* is very simple. If parties have trouble foreseeing the possible *physical* contingencies, they can write contracts that *ex ante* specify only the possible *payoff* contingencies. (After all, it is only payoffs that ultimately matter.) Then, later on, when the state of the world is realized, they can fill in the physical details. The only serious complication is incentive-compatibility: Will it be in each agent's interest to specify these details truthfully? But the techniques of the implementation literature can be used to ensure that truthful specification occurs in equilibrium.[1]

A rough analogy can be made with the use of securities in competitive markets. As Arrow (1953) showed, the competitive equilibrium of an economy with complete contingent markets can be replicated by having agents first trade in (a properly chosen set of) securities and then, after the state of nature is realized, conducting spot markets. The securities—which can be denominated in money rather than physical goods—are analogous to our payoff-denominated contracts. The spot markets correspond to "filling in the physical details."

The paper is organized as follows. Section 2 lays out the classic framework of complete contracting with describable states of nature. Section 3 defines the agents' (minimal) representation of their environment in the polar case in which states of nature are indescribable, under the maintained assumption that the agents can perform dynamic programming and so are able to envision the payoff consequences of their contract and investments. Section 4 illustrates through an example the possibility that indescribability does not constrain the payoffs that can be obtained by rational parties. Then for general environments, Section 5 establishes the irrelevance theorem (Theorem 1). It shows that in the absence of contract renegotiation the indescribability of states of nature does not interfere with optimal contracting if the optimal contract when states are describable is "welfare-neutral". A contract is welfare-neutral if, whenever two states are payoff-equivalent (*i.e.* they are distinguishable only by features not affecting the von Neumann–Morgenstern utilities), it gives rise to the same utilities in both states; welfare-neutrality is then shown to be unrestrictive under two more primitive assumptions (Section 6): a generalization of the requirement that the ratio of the agent's marginal utilities of money be independent of the state of nature (a condition which holds for instance if preferences are quasi-linear), and the condition that the relative likelihood of two payoff-equivalent states not convey information about prior unverifiable actions, if any (Theorem 2). These two assumptions are satisfied in most incomplete contract models in the literature that we are aware of.

1. Just as we do, Wernerfelt (1989) argues that contracts can be written in terms of payoffs even when future physical contingencies are unknown. The main focus of his analysis, however, is the implementation of implicit contracts through infinitely repeated games.

Hence, the irrelevance theorem applies to these models. However, we also provide examples in which the assumptions do not hold and indescribability matters. We then show that the optimal contract can be implemented in a parsimonious way, namely through the announcement of a single action on the equilibrium path, even when states of nature are indescribable (Section 7).

Section 8 examines the issue of contract renegotiation. Starting with the work of Dewatripont (1989), the economics literature has examined optimal complete contracts under the constraint that parties cannot commit not to renegotiate to their mutual advantage (see *e.g.* Fudenberg–Tirole (1990), Hart–Tirole (1988) and Maskin–Moore (1999). In this respect, it should be noted that this literature on complete contracts with renegotiation (CCR) adheres to the standard complete contracting methodology, and does not, directly at least, invoke indescribable contingencies. By contrast, the incomplete contracting approach, although usually allowing for renegotiation, seems at least motivationally distinct from this CCR literature. Indeed, the papers just mentioned are usually not associated with the incomplete contracting literature). First, we point out that renegotiation may be hard to reconcile with a framework that otherwise presumes perfect rationality. We suggest how rational parties could, in principle, commit themselves not to renegotiate. While, like transaction costs, renegotiation is pervasive in practice, we should devote more research toward reconciling its existence with current modelling practices. Nevertheless, we feel that exploring the implications of renegotiation for those models is worthwhile. Hence, we attempt to delineate when the indescribability of states of nature is costly given the possibility of renegotiation. In particular, we show that it is *not* costly in the sort of models typically considered in the literature provided that agents are at least slightly risk averse (Theorem 3).

Short of a theory of bounded rationality, one can either focus on simple institutions on *a priori* grounds or study the implications of complete contract theory in structured environments. Section 9 discusses the implications of the irrelevance results and outlines some avenues for further research.

2. THE COMPLETE CONTRACTING BENCHMARK

There are two agents ($i = 1, 2$) and three dates. [Our analysis generalizes immediately to n agents.] Date 0 is the contracting date. Date 1 is the investment date; each agent i makes an *ex post* unverifiable investment $e_i \in E_i$. Finally, at date 2, an *ex post* verifiable action a is chosen. Each pair $e = (e_1, e_2) \in E_1 \times E_2$ gives rise, stochastically, to a *verifiable* state of nature θ characterized by:

(i) a finite[2] action set A^θ,

and

(ii) payoff functions $u^\theta = (u_1^\theta, u_2^\theta)$, where $u_i^\theta : A^\theta \to \mathbb{R}$.

The assumption that payoff functions themselves are verifiable *ex post* is obviously strong. But our goal is to show that indescribability of states by itself does not reduce welfare, and this conclusion is strengthened the more we stack the deck in favour of complete contracts. Our conclusions would continue to hold *a fortiori* were we to suppose that only some *partial* signal z^θ of θ is verifiable.

2. The assumption of a finite action space for each state of nature is made for expositional and notational simplicity.

Of course, given the stochastic nature of action sets and payoff functions, parties will not in general know the state in advance, even if they know the investment levels. We assume, however, that they have prior probabilistic beliefs about states and that they share a common prior expressed by the conditional probabilities $p(\theta|e)$.[3]

Let Θ denote the set of states, and let $A \equiv \bigcup_{\theta \in \Theta} A^\theta$. A complete contract is a function

$$f: \Theta \to A \quad \text{such that } f(\theta) \in A^\theta \quad \text{for all } \theta.$$

The contract f induces an investment "game" between the players in which, given $e = (e_1, e_2)$, player i's payoff is

$$\sum_{\theta \in \Theta} p(\theta|e) u_i^\theta(f(\theta)) - c_i(e_i),$$

where $c_i(e_i)$ is agent i's cost of investment.[4]

We say that the pair (e^*, f) is *feasible* if, given f, the unique equilibrium of the investment "game" consists of each agent i selecting $e_i = e_i^*$:

(i) e^* constitutes a Nash equilibrium:

$$\sum_{\theta \in \Theta} p(\theta|e^*) u_i^\theta(f(\theta)) - c_i(e_i^*) \geqq \sum_{\theta \in \Theta} p(\theta|e_i, e_j^*) u_i^\theta(f(\theta)) - c_i(e_i) \qquad (1)$$

for $i = 1, 2$ and all $e_i \in E_i$, and

(ii) there is no other equilibrium: if we replace e^* with $e^{**} \neq e^*$, (i) no longer holds.

The complete contracting literature usually focuses on allocations that are Pareto optimal in the set of feasible allocations. When the state is verifiable, as it is here, the revelation of the state is no concern and the search for (in general second best) Pareto optimal allocations is a standard multi-person moral hazard problem.

Dividing the action space into state-dependent and state-independent components

In practice, the feasibility of some aspects of an action may be state-dependent (*e.g.* which goods are to be produced) while others (*e.g.* monetary transfers) are state-independent. Let

$$A^\theta = X^\theta \times Y$$

where X^θ is state-dependent and Y is state-independent. For concreteness, we shall assume that the choice of $y \in Y$ specifies the distribution of some private good, such as money. That is, $y = (y_1, y_2)$, where y_i is agent i's allotment of private good. Suppose that the set of feasible allocations Y of this private good is

$$Y = \{(y_1, y_2) | \underline{y}_i \leqq y_i \leqq \bar{y}_i, \ i = 1, 2 \quad \text{and} \quad y_1 + y_2 \leqq \bar{y}\}.$$

Without loss of generality take $\bar{y} = 0$. We shall suppose that u_i^θ depends on x and y_i only, and that for all $\theta \in \Theta$ and all $x \in X^\theta$, $u_i^\theta(x, y_i)$ is strictly increasing and continuous in y_i in the feasible set, labelled Y_i.

3. The common prior assumption is certainly strong, but is usually invoked in both the complete and incomplete contract literatures. Relaxing it by allowing agents to "agree to disagree" on their priors would affect the optimal complete contract but we conjecture that it would not affect the irrelevance theorem.

4. We adopt the common but strong assumption that the cost of effort is additively separable from terms involving θ because, as already noted, we are invoking the implausibly strong but expositionally useful supposition that θ is verifiable *ex post*. Thus if there were terms in the utility function involving both θ and e, it might be possible to infer e from θ, which would dispose of the moral hazard problem entailed in e not being verifiable.

Welfare-neutral complete contracts

In what follows, we will be particularly concerned with complete contracts that are *welfare-neutral*. To define this concept, let us call two states θ and θ' *equivalent* ($\theta \sim \theta'$) if $|X^\theta| = |X^{\theta'}|$ and there exists a bijection $\pi: X^\theta \to X^{\theta'}$ and scalars $a_i > 0$ and β_i such that

$$u_i^\theta(x, y_i) = a_i u_i^{\theta'}(\pi(x), y_i) + \beta_i \quad \text{for all } x \in X^\theta,\ y_i \in Y_i \quad \text{and } i = 1, 2. \tag{2}$$

That is, two states are equivalent if, up to a renaming of the state-dependent actions, the von Neumann–Morgenstern preferences are the same. A complete contract $f: \Theta \to A$ is *welfare-neutral* if, whenever θ and θ' are equivalent, we have

$$u_i^\theta(f(\theta)) = a_i u_i^{\theta'}(f(\theta')) + \beta_i \quad \text{for all } i.$$

In other words, if the von Neumann–Morgenstern utility functions are the same (modulo a positive affine transformation) in states θ and θ', then $f(\theta)$ and $f(\theta')$ give rise to the same utilities (modulo that transformation).

Section 6 will provide a detailed study of the welfare neutrality assumption. At this stage, let us content ourselves with two illustrations. First, welfare neutrality is violated in the standard *principal-agent model*. In this model, only the agent exerts effort at date 1. The state of nature θ realized at date 2 is the principal's profit or benefit from the agent's activity, and there is no decision to be taken at date 2 (X^θ is a singleton for all θ). If the principal either is risk neutral or has exponential (CARA) utility function, the principal's profit does not affect her von Neumann–Morgenstern preferences over income, and so all states are payoff equivalent. Yet, if θ is verifiable, as is assumed in the principal-agent literature, the optimal contract in general specifies a state-contingent transfer. And so welfare neutrality is not satisfied in this standard model.

In contrast, welfare neutrality is trivially satisfied in the two-sided-investment *buyer-seller model*, which has been one of the main applications of the incomplete contracting literature. In the complete contract counterpart to this model, the buyer's and the seller's date-1 investments in human capital, say, are denoted e_B and e_S, respectively. The buyer and the seller trade one unit of a single good at date 2 (by convention, "no trade" corresponds to a trade of a useless and costless good). There are n goods $k = 1, \ldots, n$, generating value $v_B^k(\theta)$ for the buyer and cost $c_S^k(\theta)$ for the seller. So, the decision set is $X = \{1, \ldots, n\}$. The utilities if good k is traded are

$$u_B[v_B^k(\theta) + y_B] - e_B \quad \text{and} \quad u_S[-c_S^k(\theta) + y_S] - e_S,$$

respectively. [Applications of this model to property rights usually also assume that the parties have outside opportunities, in that the seller can sell to alternative buyers instead, and the buyer similarly can buy from alternative sellers instead. These outside opportunities do not affect the fact that the optimal complete contract is welfare neutral.] The incomplete contracting literature usually further assumes that e_B induces a probability distribution over ordered valuation vector $\{v_B(1), v_B(2), \ldots, v_B(n)\}$ where $v_B(k)$ is the kth statistic, and that the ordered valuation vector is a sufficient statistic for θ to learn e_B; and similarly for the seller. In these circumstances, the optimal complete contract (can be chosen so as to) yield the same valuation, cost and transfers in two states θ and θ' that have the same ordered valuation and cost vectors. So welfare neutrality holds.

3. INDESCRIBABLE STATES

As we have laid out the model of Section 2, the complete contract f is chosen to solve a pure moral hazard problem as formulated in (1). If we drop the assumption that u^θ is

verifiable *ex post* but keep the assumption that X^θ is describable and verifiable, the model is in the spirit of the contributions of Hart and Moore (1988), Aghion *et al.* (1994) and Nöldeke and Schmidt (1995), among others.[5] In this case, however, $f(\theta)$ can no longer be directly implemented because the action it specifies depends on unverifiable information. The agents have to play a *mechanism* (see below) in order to determine the appropriate action a. More generally, we want to know what can be achieved not only when u^θ is unverifiable, but also when the states of nature are indescribable.

Since the goal of this paper is to assess inefficiencies resulting from the indescribability of actions and contingencies *ex ante*, we must compare two payoff-equivalent representations of the world: one in which the agents can perfectly describe the actions and contingencies *ex ante*, and another in which they cannot. Either way we follow the incomplete contract literature by assuming that both agents learn θ *ex post*,[6] and that whether or not a given action x is feasible can be verified *ex post*.

If parties are able to perform dynamic programming (which incomplete contract models always assume), then at the very least they can formulate a probability distribution over the possible payoffs. To represent possible payoffs, let

$$v_i: \{1, \ldots, m\} \times Y_i \to \mathbb{R},$$

be a "dummy" or "number-based" payoff function for agent i, where $\{1, \ldots, m\}$ is a set of positive integers. The only difference between v_i and an ordinary (*action-based*) payoff funtion u_i is that, in the former, the domain of physical actions has been replaced by this set of integers. The integers are a dummy index reflecting the fact that agents cannot forecast physical actions.[7]

Viewed this way, a state $\theta = (X^\theta, u^\theta) \in \Theta$ can alternatively be expressed as $(X^\theta, v^\theta, \sigma^\theta)$, where:

(i) X^θ is the (physical) action space in state θ, where $|X^\theta| = m^\theta$;

(ii) $v^\theta = (v_1^\theta, v_2^\theta)$, where $v_i^\theta: M^\theta \times Y_i \to \mathbb{R}$ constitutes agent i's number-based payoff function and $M^\theta = \{1, \ldots, m^\theta\}$;
and

(iii) $\sigma^\theta: M^\theta \to X^\theta$ is a state-θ mapping from numbers into actions in X^θ. That is, $\sigma^\theta(k)$ is a physical action that implements the payoff pair $v^\theta(k, y) = (v_1^\theta(k, y_1), v_2^\theta(k, y_2))$ in state θ: $u_i^\theta(\sigma^\theta(k), y_i) = v_i^\theta(k, y_i)$ for $i = 1, 2$. The function σ^θ can be interpreted as a *deciphering key*.

We let V denote the set of *ex ante* possible number-based payoff functions pairs $v = (v_1, v_2)$.

When we say that agents can perform dynamic programming in the case of indescribable states, we mean that they can formulate prior beliefs $\hat{p}(v|e)$ about how investments affect payoffs. Beliefs $\hat{p}(\cdot|\cdot)$ are consistent with beliefs in the describable states model if, for all $v \in V$ and all $e \in E$,

$$\hat{p}(v|e) = \sum_{\theta \in \Theta_v} p(\theta|e), \tag{3}$$

5. One interpretation of this modified model (where u^θ is unverifiable) is that, for each θ, agents can describe the possible actions X^θ well enough in advance so that these can be specified in the contract (in the articles mentioned above, agents know *ex ante* which good they should trade), but that they are uncertain *ex ante* about how much utility the actions bring.

6. This rules out *ex post* asymmetric information. We conjecture, however, that the irrelevance theorem extends the case of *ex post* asymmetric information in the sense that the *ex ante* describability or indescribability of states continues not to matter under our assumptions.

7. Jean-Charles Rochet has suggested that there is a rough analogy between this construction and Kolmogorov's approach to probabilities (in probability theory, we need not know the physical properties of states of nature, since we care only about the probability distribution over consequences.)

where

$$\Theta_v = \{\theta \in \Theta \mid v^\theta = v\}.$$

Number-based contracts

When actions are no longer describable in advance, parties cannot pre-specify the action they would like to implement in each state. However, they can in principle pre-specify the *utilities* they would like to implement. Call

$$\hat{f}: V \to Z_+ \times Y,$$

a *number-based contract*, where for any number-based payoff functions $v = (v_1, v_2)$, $\hat{f}(v)$ specifies an integer k and a describable action $y \in Y$ such that $(v_1(k, y_1), v_2(k, y_2))$ are the corresponding utilities to be implemented.

By analogy with the definition for action-based contracts, a number-based contract $\hat{f}$ is *welfare-neutral* if, whenever $\theta \sim \theta'$ (so that there exist a one-to-one mapping $\pi: M^\theta \to M^{\theta'} = M^\theta$ and scalars $a_i > 0$ and β_i for $i = 1, 2$ such that $v_i^\theta(k, y_i) = a_i v_i^{\theta'}(\pi(k), y_i) + \beta_i$ for $i = 1, 2$ and all (k, y_i)), then

$$v_i^\theta(k^\theta, y_i^\theta) = a_i v_i^{\theta'}(k^{\theta'}, y_i^{\theta'}) + \beta_i \quad \text{for } i = 1, 2,$$

where $\hat{f}(v_1^\theta, v_2^\theta) = (k^\theta, y^\theta)$ and $\hat{f}(v_1^{\theta'}, v_2^{\theta'}) = (k^{\theta'}, y^{\theta'})$.

The number-based contract $\hat{f}$ *corresponds* to the complete contract $f: \Theta \to A$, if for all $v \in V$ and for all $\theta \in \Theta$ such that $v^\theta = v$,

$$v_i(\hat{f}(v)) = u_i^\theta(f(\theta)) \quad \text{for } i = 1, 2.$$

It is not hard to see that a complete contract f is welfare-neutral if, and only if, there exists a welfare-neutral number-based contract corresponding to f.

Mechanisms

There are two difficulties with ensuring that the utilities prescribed by number-based contract $\hat{f}$ actually get implemented: (i) the true number-based payoff functions (v_1^θ, v_2^θ) are not verifiable and (ii) the true deciphering key σ^θ, mapping integers to physical actions, is not verifiable. Accordingly, $\hat{f}$ must be implemented indirectly; the agents' contract must specify a mechanism for them to play. A *mechanism* (or game form) g is a procedure in which, once θ is realized, one or both agents propose a feasible action or actions as part of their strategies (strategies may involve more than just proposing actions, *e.g.* making claims about the true utility functions). On the basis of the agents' strategies, g determines a physical action. In view of the *ex ante* indescribability of the action space, however, the rule mapping strategies to actions cannot depend on any physical property of these actions. (Thus, the rule "implement the action that agent 1 has proposed" is permissible, but the rule "implement the action involving production of the higher quality good" is not.) In any state $\theta \in \Theta$, g induces a game between the two agents—each agent i chooses a strategy to maximize the expected value of u_i^θ.

We shall say that g with joint strategy space $S_1 \times S_2$ *implements* the number-based contract $\hat{f}$ if, for all $\theta \in \Theta$ and all (subgame-perfect) equilibria $(s_1^\theta, s_2^\theta)(\in S_1 \times S_2)$ in state θ,

$$u_i^\theta(g(s_1^\theta, s_2^\theta)) = v_i^\theta(\hat{f}(v_1^\theta, v_2^\theta)) \quad \text{for } i = 1, 2.$$

That is, the equilibrium utilities[8] in state θ are the same as those prescribed by $\hat{f}$ for agents' number-based payoff functions in state θ.

Our "irrelevance theorem" provides conditions under which a number-based contract $\hat{f}$ is implementable. A necessary requirement is that $\hat{f}$ be welfare-neutral. To see this, note that if $\theta \sim \theta'$ then the game induced by an implementing mechanism after θ is strategically equivalent to that induced after θ'. Hence, if (μ_1, μ_2) is an equilibrium payoff for the latter game, $(a_1\mu_1 + \beta_1, a_2\mu_2 + \beta_2)$ is an equilibrium payoff for the former, *i.e.* $\hat{f}$ is welfare-neutral.

Before turning to this result, we first examine the simpler special case in which unlimited monetary transfers can be made and payoff functions are quasi-linear.[9]

4. AN EXAMPLE OF IMPLEMENTATION WITH LARGE PENALTIES

Assume that:

(a) for all $\theta \in \Theta$, utility functions are quasi-linear[10] between actions in X^θ and Y

$$u_i^\theta(x, y_i) = w_i^\theta(x) + y_i, \qquad i = 1, 2;$$

(b) there is no bound on the magnitude of individual agents' transfers

$$Y = \{(y_1, y_2) | y_1 + y_2 \leqq 0\};$$

(c) there exists a describable action x^0 (which can be interpreted as the "no-trade" point) such that, for all $\theta \in \Theta$, $x^0 \in X^\theta$. We shall normalize the payoff from x^0 to zero

$$w_i^\theta(x^0) = 0 \quad \text{for all } \theta \in \Theta, \qquad i = 1, 2;$$

(d) $\hat{f}$ is a Pareto-optimal number-based contract.

Consider the following mechanism:

Stage 1: Agent 1 announces:
— an action set X;
— number-based payoff functions v_i: $M \to \mathbb{R}$, $i = 1, 2$, where $|M| = |X|$ and $(v_1, v_2) \in V$; and
— a deciphering key σ: $M \to X$.

(If agent 1 fails to do this, the outcome is $(x^0, (-P, P))$, where P is a large monetary payment; *i.e.* action x^0 is implemented and agent 1 pays agent 2 a penalty P). Note that,

8. In the implementation literature, a contract—or, more generally, a *social choice rule*—is said to be implemented by a mechanism if, for each state $\theta \in \Theta$, the equilibrium *physical outcome* of the mechanism is the same as that prescribed by the contract. When physical outcomes cannot be described in advance, however, the best we can do is to ensure that equilibrium *utilities* are the same as those prescribed by the contract.

9. A well-known but highly special illustration of the idea that details of a contract may be filled in *ex post* is the following. A seller is to sell one unit of a good to a buyer. The unit has value v if the good is the "relevant good" and 0 otherwise. What constitutes the relevant good is *ex ante* unknown, and perhaps even the possibilities are "indescribable". Yet there exists an efficient contract as follows: (i) the seller describes a good *ex post*, (ii) the buyer accepts or rejects the good. If she accepts, she must pay v; if she rejects, she pays 0. In the absence of renegotiation (an issue to be discussed later), this contract achieves the efficient outcome, and gains from trade can be divided *ex ante* through an arbitrary transfer between the two parties.

10. The assumption that $u_i^\theta(x, y_i)$ is *linear* in y_i simplifies but is not crucial to the following construction. A nonlinear function $q_i(y_i)$ would do as well. The key property of assumption (*a*) used in the construction below is that u_i^θ is separable in x and y_i.

because utility is separable, we can ignore the utility from money and view $v_i(k)$ as an announcement of the *gross* surplus $w_i^{\theta}(\sigma(k))$.

Stage 2: Agent 2 decides to challenge or not to challenge agent 1's announcement.
— If he does not challenge, the outcome is (x^*, y^*), where $x^* = \sigma(k^*)$ and $(k^*, y^*) = \hat{f}(v_1, v_2)$.
— If he does challenge, then agent 1 must pay a large penalty P to agent 2.

Agent 2 can challenge in one of three ways:

(i) Agent 2 can challenge the action space X either by exhibiting a feasible action $x \notin X$ or by exhibiting an infeasible $x \in X$. In either case, the implemented action is x^0. That is, the outcome is

$$(x^0, (-P, P));$$

(ii) Agent 2 can challenge agent 1's implicit claim that $v_1(\sigma^{-1}(\cdot))$ is agent 1's actual surplus function. Namely, he chooses integer k and a number $\varepsilon \neq 0$, and gives agent 1 the choice between options (a) and (b), where (a) consists of action $\sigma(k)$ and agent 1 receiving no monetary transfer (except for the transfer $-P$ corresponding to the penalty he paid) and (b) consists of the action x^0 and a transfer of $v_1(k) + \varepsilon$ (beyond the penalty $-P$). Agent 1 then chooses between (a) and (b). The challenge is successful if either $\{\varepsilon > 0$ and agent 1 picks option (a)$\}$ or $\{\varepsilon < 0$ and agent 1 picks option (b)$\}$. Otherwise, it is unsuccessful. In the case of an unsuccessful challenge, agent 2 must himself pay a penalty $2P$, and this goes to a third party. To summarize, if agent 1 selects (a), the outcome is

$(\sigma(k), (-P, P))$ if the challenge is successful $(\varepsilon > 0)$,

$(\sigma(k), (-P, -P))$ if the challenge is unsuccessful $(\varepsilon < 0)$.
If agent 1 selects (b), the outcome is

$(x^0, (v_1(k) + \varepsilon - P, P - v_1(k) - \varepsilon))$ if the challenge is successful $(\varepsilon < 0)$, and

$(x^0, (v_1(k) + \varepsilon - P, -P))$ if the challenge is unsuccessful $(\varepsilon > 0)$;

(iii) Agent 2 can challenge agent 1's claim that $v_2(\sigma^{-1}(\cdot)) = w_2(\cdot)$ is agent 2's actual utility function. Specifically, agent 2 can announce a number-based payoff function v_2'. In this case agent 2 pays P (the penalty received from agent 1) to a third party. Agent 1 then has the opportunity to counter-challenge. If he refrains from doing so, the alternative (x^{**}, y^{**}) is implemented, where $x^{**} = \sigma(k^{**})$, $(k^{**}, y^{**}) = \hat{f}(v_1, v_2')$. If agent 1 does counter challenge, then agent 2 must pay a second penalty P to agent 1 (implying that agent 2's overall net payment is P) and challenge (ii) is carried out with the roles reversed.

We claim that the mechanism implements $\hat{f}$ if P is big enough. Challenge (i) exploits the fact that a court can verify whether a specific action is feasible or not. This challenge thus guarantees that agent 1 will not misstate the true action space. Challenge (ii) guarantees that agent 1 (and the counter-challenge in (iii) guarantees that agent 2) will not lie about his own payoff function. Finally, challenge (iii) ensures that agent 1 will not lie about agent 2's payoff function in a way that reduces the latter's payoff beyond that prescribed by $\hat{f}$. Given that $\hat{f}$ is Pareto optimal, agent 1 therefore cannot obtain more than the payoff that $\hat{f}$ specifies.

5. THE IRRELEVANCE THEOREM

Let us drop the assumptions of Section 4 that payoff functions are separable and that unbounded monetary transfers are possible. Instead let us assume that, given the number-based contract $\hat{f}$, the following two technical assumptions are satisfied:

Assumption 1 There exist a describable alternative x^0 and transfers $y^0 \in Y((x^0, y^0)$ could be a "no-trade" point) such that, for all $\theta \in \Theta$,

(i) $x^0 \in X^\theta$,
(ii) $v_i^\theta(\hat{f}(v^\theta)) > u_i^\theta(x^0, y_i^0)$, and
(iii) $\underline{y}_i < y_i^0 < \bar{y}_i$, for $i = 1, 2$

and

Assumption 2. For all $\theta \in \Theta$, if $\hat{f}(v^\theta) = (k^\theta, y^\theta)$ then $\underline{y}_i < y_i^\theta < \bar{y}_i$ for $i = 1, 2$.

Remark. Assumptions 1 and 2 are usually satisfied in economic models. In some models they may hold with only weak inequalities, as is the case when an agent i is protected by limited liability but, because of the moral hazard problem, it is desirable to "punish" him by assigning the lowest possible transfer $\underline{y}_i$. In such a case, Theorem 1 below says that one can approximate the outcome with describable actions arbitrarily closely even if actions are indescribable. Hence, Assumptions 1 and 2 are not very restrictive.

Theorem 1. *If the number-based contract $\hat{f}$ corresponds to a complete contract f that is welfare-neutral, Pareto-optimal, and satisfies Assumptions* 1 *and* 2,[11] *then $\hat{f}$ can be implemented in subgame-perfect equilibrium when states are indescribable.*

Remark. The proof of Theorem 1 builds on standard implementation theory. In the implementation literature (see Moore (1992) and Palfrey (1998) for recent surveys), the objective is to construct mechanisms that elicit agents' payoffs when the action space is known in advance but payoffs are neither known *ex ante* nor verifiable *ex post*. We show that such mechanisms can be extended to the case in which the action space cannot be forecast. Specifically, after the state θ is realized, we have agent 1 announce what he claims are (i) the realized action space X and (ii) a mechanism g^X that, given X, implements the number-based contract $\hat{f}$ (the existence of an implementing mechanism is assured by the implementation literature). We then allow agent 2 to challenge either announcement. Because we assume (as does the incomplete contract literature) that courts can verify *ex post* whether or not any given action is feasible, it is easy for agent 2 to challenge successfully agent 1 if he has omitted a feasible action from X or included an infeasible one. (In the contract we construct here, agent 1 must announce the entire feasible action space. In Section 7, however we will show that the parties can make do with much more parsimonious announcements). Moreover, assuming that X is the true feasible set, agent 2 can

11. We could drop the stipulation that $\underline{y}_i < y_i^\theta$ in Assumption 2 and also the hypothesis that f is Pareto optimal if instead we assumed:

Assumption 3. For all $\theta \in \Theta$ and all $i = 1, 2$, if $(\hat{x}, \hat{y})$ satisfies $u_i^\theta(\hat{x}, \hat{y}) \geq u_i^\theta(x, y)$ for all (x, y) such that $u_j^\theta(x, y) \geq u_j^\theta(x^0, y^0)$, where $j \neq i$, then $v_j^\theta(\hat{f}(v^\theta)) > u_j^\theta(\hat{x}, \hat{y})$.

In words, if agent i chooses his favorite alternative subject to agent j getting at least the utility of (x^0, y^0), then agent j's utility is strictly less than that from the number-based contract $\hat{f}$.

In fact, Assumption 3 is implied by Assumption 2 and Pareto optimality of $\hat{f}$.

mechanically prove when a mechanism g^X fails to implement $\hat{f}$ by simply exhibiting a nonoptimal equilibrium (which also constitutes a successful challenge). The incentive for agent 2 to challenge successfully is that she is then rewarded with the opportunity to pick her favourite action (this deters agent 1 from announcing a false action space or non-implementing mechanism). If agent 2 does not challenge—which will occur only if agent 1's claim is true—the mechanism g^X is then played.

Proof. Note first that, from part (iii) of Assumption 1, there exists $y^{00} \in Y$ such that $y_i^{00} < y_i^0$, $i = 1, 2$. Hence,

$$u_i^\theta(x^0, y_i^{00}) < u_i^\theta(x^0, y_i^0) \quad \text{for all } \theta \in \Theta, \qquad i = 1, 2.$$

Given action space X, we shall call a mechanism g^X, with joint strategy space $S_1^X \times S_2^X$, *successful* provided that, for all possible pairs of number-based payoff functions $v \in V$ (where $v_i: M \times Y_i \to \mathbb{R}$, $i = 1, 2$) for which $|X| = |M|$ and all possible deciphering keys $\sigma: M \to X$, the unique subgame-perfect equilibrium payoffs of the game g^X when agents have utility functions $(v_1(\sigma^{-1}(\cdot), \cdot), v_2(\sigma^{-1}(\cdot), \cdot))$ are $v(\hat{f}(v))$.

For the time being, suppose that, for all action spaces that could arise—*i.e.* for all X for which there exists $\theta \in \Theta$ with $X^\theta = X$—there exists a successful mechanism. (We will show this is so below. This is where we make use of implementation theory.) Then the parties can sign a contract that stipulates that, once the state of nature is realized:

(i) agent 1 proposes a feasible set of actions X and a corresponding mechanism g^X;
(ii) agent 2 can accept the proposal or challenge it;
(iii) if agent 2 challenges, then he can (a) exhibit a feasible action $x \notin X$, or (b) exhibit an infeasible action $x \in X$, or (c) demonstrate that, given X, the proposed mechanism g^X is not successful;
(iv) if agent 2 succeeds with any of (a)–(c) then he can name an action (x, y), which is implemented; if he fails with the challenge the outcome is (x^0, y^{00}); in either case, the execution of the contract ends at this point;
(v) if agent 2 accepts agent 1's proposal, then the mechanism g^X is played and the action that this leads to is implemented, provided it is feasible (if not, the outcome is (x^0, y^{00})).

We claim that, for all $\theta \in \Theta$, the unique equilibrium payoffs of this contract are $v^\theta(\hat{f}(v^\theta))$. To see this, note that if agent 1 proposes the true feasible set X^θ and a corresponding successful mechanism g^{X^θ} and if agent 2 accepts, then, by definition of "successful mechanism," the resulting continuation equilibrium payoffs are indeed $v^\theta(\hat{f}(v^\theta))$. Moreover, it is (uniquely) optimal for agent 2 to accept this proposal, since any challenge would fail and therefore result in outcome (x^0, y^{00}), which from Assumption 1 (ii) is worse than $v_2^\theta(\hat{f}(v^\theta))$. It remains only to show that in equilibrium agent 1 must make such a proposal. Observe that if he did not make such a proposal, then agent 2 could challenge successfully (either by showing that X is not the true feasible set or by exhibiting number-based payoff functions $(v_1, v_2) \in V$, a deciphering key σ, and an equilibrium of g^X when agents have utility functions $(v_1(\sigma^{-1}(\cdot), \cdot), v_2(\sigma^{-1}(\cdot), \cdot))$ such that agents' equilibrium payoffs are not $v(\hat{f}(v))$. Moreover, since $\underline{y}_1 < y_1^\theta$ and $y_2^\theta < \bar{y}_2$, by doing so, agent 2 could get a strictly higher payoff than $v_2^\theta(\hat{f}(v^\theta))$. But because f is Pareto optimal, this would imply a payoff strictly less than $v_1^\theta(\hat{f}(v^\theta))$ for agent 1, a suboptimal outcome.

To complete the proof, we must show that, for all $\theta \in \Theta$, there exists a successful mechanism g^{X^θ}. To do this, it is enough to show that the sufficient conditions in Moore and Repullo (1988) are satisfied. We relegate this portion of the proof to Appendix 1. ||

6. HOW RESTRICTIVE IS WELFARE-NEUTRALITY?

The crucial hypothesis in Theorem 1 is the welfare-neutrality of the contract f. We next show (Theorem 2) that this hypothesis is unrestrictive provided that two more fundamental conditions hold: (i) state-independence of the ratios of the marginal utilities of money and (ii) unidentified effort.

Definition. Preferences satisfy the property of state-independence of the ratios of the marginal utilities of money if, for $\theta, \theta' \in \Theta$, whenever θ and θ' are equivalent—*i.e.* there exist one-to-one mapping $\pi: X^{\theta} \to X^{\theta'}$ and scalars $a_i > 0$, β_i such that (2) holds—then

$$a_1 = a_2.$$

Observe that a sufficient condition for state-independence is that payoff functions take a separable form:

$$u_i^{\theta}(x, y_i) = w_i^{\theta}(x) + a(\theta)q_i(y_i). \tag{4}$$

Definition. Effort is unidentified if whenever θ and θ' are equivalent

$$p(\theta|e, \{\theta, \theta'\}) = p(\theta|e', \{\theta, \theta'\}) \quad \text{for all } e \text{ and } e'.$$

That is, effort is unidentified when the only information conveyed by the realized state about the effort levels chosen is that inherent in the von Neumann–Morgenstern preferences.

Theorem 2. *Suppose that the preferences satisfy the property of state-independence of the ratios of the marginal utilities of money and effort is unidentified. Then for any complete contract f, there exists a welfare-neutral complete contract f' such that*

$$\sum_{\theta \in \Theta} p(\theta|e)u^{\theta}(f(\theta)) = \sum_{\theta \in \Theta} p(\theta|e)u^{\theta}(f'(\theta)) \quad \text{for all } e.$$

Proof. Consider two equivalent states θ and θ'. From state-independence there exist $\pi: X^{\theta} \to X^{\theta'}$ and scalars $a > 0$, β_i such that

$$u_i^{\theta}(x, y_i) = au_i^{\theta'}(\pi(x), y_i) + \beta_i \text{ for all } (x, y_i) \in X^{\theta} \times Y_i \text{ and } i = 1, 2. \tag{5}$$

Now, conditional on e and $\{\theta, \theta'\}$, agent i's expected utility from the complete contract is

$$pu_i^{\theta}(x^{\theta}, y_i^{\theta}) + (1-p)u_i^{\theta'}(x^{\theta'}, y_i^{\theta'}), \tag{6}$$

where $f(\theta) = (x^{\theta}, y^{\theta})$ and $f(\theta') = (x^{\theta'}, y^{\theta'})$ and

$$p = p(\theta|e, \{\theta, \theta'\}). \tag{7}$$

Thanks to (5) we can rewrite (6) as:

$$pau_i^{\theta'}(\pi(x^{\theta}), y_i^{\theta}) + (1-p)u_i^{\theta'}(x^{\theta'}, y_i^{\theta'}) + p\beta_i. \tag{8}$$

Now consider a contract f' in which, when θ occurs, there is a $[t, 1-t]$ randomization between (x^{θ}, y^{θ}) and $(\pi^{-1}(x^{\theta'}), y^{\theta'})$. And when θ' occurs, there is a $[t, 1-t]$ randomization between $(\pi(x^{\theta}), y^{\theta})$ and $(x^{\theta'}, y^{\theta'})$. Notice that this is welfare-neutral. Moreover, it results

in conditional expected utility

$$p(tu_i^{\theta}(x^{\theta}, y_i^{\theta}) + (1-t)u_i^{\theta}(\pi^{-1}(x^{\theta'}), y_i^{\theta'})) + (1-p)(tu_i^{\theta'}(\pi(x^{\theta}), y_i^{\theta}) + (1-t)u_i^{\theta'}(x^{\theta'}, y_i^{\theta'})),$$

which, from (5) can be rewritten as:

$$(apt + (1-p)t)u_i^{\theta'}(\pi(x^{\theta}), y_i^{\theta}) + (ap(1-t) + (1-p)(1-t))u_i^{\theta'}(x^{\theta'}, y_i^{\theta'}) + p\beta_i. \tag{9}$$

Take

$$t = \frac{ap}{ap + 1 - p}.$$

This is well-defined since, from unidentified effort, p does not depend on e. With this substitution (9) reduces to (8). That is, we can replicate the expected utility from f—conditional on e, θ and θ'—using the welfare-neutral contract f'. Continuing in the same way, we can establish the same thing for all equivalent pairs θ and θ'. ||

To understand the role of state-independence of the ratios of the marginal utilities of money in Theorem 2, it may be helpful to consider an example in which it is violated and Theorem 1 does not hold. There are two actions a_1 and a_2 and two equally likely states of nature θ and θ'. The utility functions are

$$u_1^{\theta}(a) = \begin{cases} -e^{2\delta}, & a = a_1, \\ -1, & a = a_2, \end{cases} \qquad u_2^{\theta}(a) = \begin{cases} -e^{-2\delta}, & a = a_1, \\ -1, & a = a_2, \end{cases}$$

and

$$u_1^{\theta'}(a) = \begin{cases} -e^{-2\delta}, & a = a_1, \\ -1, & a = a_2, \end{cases} \qquad u_2^{\theta'}(a) = \begin{cases} -e^{2\delta}, & a = a_1, \\ -1, & a = a_2. \end{cases}$$

[The interpretation is that the two agents have exponential utility functions $-e^{-c_i}$, and face endowment shocks $(-\delta, \delta)$ in state θ, and $(\delta, -\delta)$ in state θ'. Action a_1 is a "gambling action," which doubles the individual shocks, and a_2 is a "hedging action."] Note that, if π denotes the permutation in which $\pi(a_1) = a_2$ and $\pi(a_2) = a_1$, then, for all a,

$$u_1^{\theta}(a) = e^{2\delta}u_1^{\theta'}(\pi(a))$$

and

$$u_2^{\theta}(a) = e^{-2\delta}u_2^{\theta'}(\pi(a)).$$

Hence states θ and θ' are equivalent, but state-independence is violated.

Consider the complete contract f for which $f(\theta) = f(\theta') = a_2$. This is an efficient contract since a_2 provides perfect insurance, *i.e.* it results in expected payoffs $(-1, -1)$. Note, however, that it is not welfare-neutral and therefore cannot be implemented in the indescribable actions case. Indeed, the only expected payoff pairs that are possible in that case are

$$\{\lambda[-\tfrac{1}{2}(1 + e^{-2\delta})] + (1-\lambda)[-\tfrac{1}{2}(1 + e^{2\delta})], (1-\lambda)[-\tfrac{1}{2}(1 + e^{-2\delta})] + \lambda[-\tfrac{1}{2}(1 + e^{2\delta})]\},$$

where $\lambda \in [0, 1]$. This follows because the endowment shocks are payoff-irrelevant at the implementation stage and therefore cannot be elicited, implying that for $i = 1, 2$, the probability that the action preferred by player i is chosen by the complete contract is the same in both states of nature.

Next, let us turn to an example illustrating the role of unidentified effort. There are two states of nature, θ and θ', and a pair of actions, $\{a_1, a_2\}$ and $\{a'_1, a'_2\}$, in each state respectively. The payoff functions are

$$u_1^{\theta}(a) = \begin{cases} 3, & a = a_1, \\ 5, & a = a_2, \end{cases} \qquad u_2^{\theta}(a) = \begin{cases} 1, & a = a_1, \\ 4, & a = a_2, \end{cases}$$

and

$$u_1^{\theta'}(a) = \begin{cases} 2, & a = a'_1, \\ 0, & a = a'_2, \end{cases} \qquad u_2^{\theta'}(a) = \begin{cases} 5, & a = a'_1, \\ 2, & a = a'_2. \end{cases}$$

Note that if π is the mapping $a'_2 = \pi(a_1)$ and $a'_1 = \pi(a_2)$

$$u_1^{\theta}(a) = u_1^{\theta'}(\pi(a)) + 3,$$

and

$$u_2^{\theta}(a) = u_2^{\theta'}(\pi(a)) - 1,$$

for $a = a_1, a_2$. Hence θ and θ' are equivalent and state-independence of the ratios of the marginal utilities of money holds. At date 1, agent 2 "works" or "shirks." Working (which costs agent 2 one util) results in state θ and shirking (which has no cost) in state θ'. So the unidentified effort condition is violated. With a complete contract it is straightforward to obtain the overall payoffs (5, 4): It suffices to specify that actions a_2 and a'_2 will be chosen in states θ and θ' respectively. This induces agent 2 to work, as $4 > 2$. However, in the case of indescribable states, if a_2 is implemented in state θ then $\pi(a_2) = a'_1$ must be implemented in state θ'. Hence, agent 2 cannot be induced to work, and the only possible *ex ante* utilities are (2, 5) or (0, 2) or mixtures thereof. This is because effort is payoff-irrelevant at date 2 and the probability of action a ($a = a_1, a_2$) in state θ must therefore be equal to the probability of action $\pi(a)$ in state θ'. Hence, if λ is the probability of action a_1 in state θ, agent 2 obtains net utility

$$\lambda + 4(1 - \lambda), \quad \text{if he works,}$$

and

$$2\lambda + 5(1 - \lambda), \quad \text{if he shirks.}$$

The agent therefore always shirks when the actions are indescribable.

We defined "complete contracting" in Section 2 as the case in which all aspects of a state—including payoffs—are verifiable *ex post*. Our work/shirk example illustrates one situation in which payoffs are verifiable indirectly. Specifically, note that the action sets are different in the two states and that there is a one-to-one correspondence between action sets and utility functions. Thus it need not literally be true that utility functions can be inspected for them to be verifiable; it may suffice that action spaces be verifiable.

Finally, let us consider a slightly different example of unidentified effort. Suppose that at date 1 agent 2 can either work or shirk to produce a widget, which is consumed by agent 1. The widget can turn out to be of either high or low quality, but agent 2 working raises the probability of a high-quality widget. Let us suppose that if agent 2 works, the widget, whatever its quality, has a blue colour, whereas if he shirks it is red. Color is entirely payoff-irrelevant. Nevertheless it ensures that in a complete contract setting, agent 2 can be perfectly insured (let us suppose that he is risk averse): The contract can stipulate that agent 1 pay agent 2 if, and only if, the widget is blue. In the case where

colors cannot be foreseen, however, agent 2 must bear some risk: for proper incentives, the payment he receives must be strictly monotonic in the quality of the widget. Moreover, this is true whether or not quality itself is verifiable. (If quality is not verifiable, then there must be some probability of no trade in the lower quality state to implement such a monotonic schedule.)

We have provided three illustrations of how Theorems 1 and 2 can break down. The assumption of welfare neutrality is by no means innocuous. Indeed, complete contract theory often focuses on situations in which welfare neutrality is violated; for example, and as we already observed, in moral hazard models, the agent's reward depends on signals of performance, such as the principal's profit, that once realized, have no impact on the parties' von Neumann–Morgenstern preferences. Such measures of performance could not be elicited from the parties *ex post*, and so they have to be describable *ex ante*.

We must point out, however, that the features driving these examples in which describability matters have not been invoked by the incomplete contract literature. Indeed, virtually all the models from that literature of which we are aware satisfy the hypotheses of Theorem 2, so that the irrelevance theorem applies.

7. PARSIMONY

The mechanisms described in Sections 4 and 5 are demanding in the sense that agents are required to fully specify the realized state of nature *ex post*. Although this is presumably a less arduous task than describing the entire space Θ *ex ante*, it may still be formidable.

Fortunately, there is a simple modification of these mechanisms that, in equilibrium, requires agents to do nothing more than describe a single action. Specifically, suppose (for expositional simplicity only) that preferences are quasi-linear. Let us add two preliminary stages to the mechanism described in Section 4. In the first stage, agent 1 proposes an action (x, y). In stage 2, agent 2 accepts or rejects the proposal. If he accepts, the outcome is (x, y), provided that this is feasible. (If it is infeasible, the "fall-back" option (x^0, y^0) is implemented.) If he rejects, then agent 2 pays an (arbitrarily small) fee ε to a third party and the mechanism constructed in Section 4 is played. Thus, we have introduced a preliminary round in which agent 1 can avoid the (possibly costly) enumeration of the action space by offering to settle on a particular action. The fee ε paid by agent 2 is invoked only to prevent frivolous challenges by that agent.

This mechanism implements the complete contract outcome arbitrarily closely (as ε goes to 0), and yet elicits only a single action on the equilibrium path: If $\hat{f}$, the number-based contract to be implemented, is Pareto-optimal, then, in any state θ, it is an equilibrium of this modified mechanism for agent 1 to propose an alternative a^θ that is ε-optimal and for agent 2 to accept. To see this, note that agent 1 cannot get away with proposing an alternative a that yields him a payoff more than $v_1^\theta(\hat{f}(v^\theta)) + \varepsilon$, since, given $\hat{f}$'s Pareto-optimality, player 2 would necessarily get less than $v_2^\theta(\hat{f}(v^\theta)) - \varepsilon$. That is, agent 2 would be better off rejecting a, since, by the Section 4 construction, the equilibrium of the continuation game gives him utility $v_2^\theta(\hat{f}(v^\theta))$. Any other equilibrium of the modified mechanism must yield the same payoffs.

We have observed that the mechanism can be chosen such that, in equilibrium, agents describe a single action. To show this, we assumed that describing the entire realized action space X^θ is feasible, although possibly very costly. We now show that a parsimony result obtains in the case in which describing the realized action space is infeasible. Assume that utilities are quasi-linear and that, as in Section 4, punishments can be large (our first parsimony result does not rely on this assumption and so, strictly speaking the second

result is not a generalization of the first; however, we believe that quasi-linearity can be dropped in the second case as well.) Let us remind the reader of our convention that, in the quasi-linear case, $v_i(k)$ is a *gross* surplus.

Consider the following mechanism: At stage 1, agent 1 announces (i) number-based payoff function (v_1, v_2), (ii) an action x, and (iii) the deciphering key $x = \sigma(k)$ for this action only. [So, player 1 announces only a single physical action.] The implicit claim in this announcement is that (v_1, v_2) are the true payoff functions and that $x = \sigma(k)$, where $(k, y) = \hat{f}(v_1, v_2)$ and σ is the true deciphering key. At stage 2, agent 2 can (i) accept agent 1's announcement, in which case (x, y) is implemented; (ii) challenge agent 1, in which case, as in the mechanism of Section 4, agent 1 must pay 2 a large fine P; or (iii) query agent 1. If agent 2 challenges, he is, in effect, denying at least one of agent 1's claims that (a) x is feasible, (b) $(v_1(\sigma^{-1}(\cdot)), v_2(\sigma^{-1}(\cdot)))$ are the true gross surplus functions and (c) $x = \sigma(k)$. How the challenge (and possible counter-challenge) is conducted and the consequences thereof are exactly as in Section 4. If agent 2 queries, then he must pay a small fee ε and can either (a) name an integer l and ask agent 1 to specify an action $\sigma(l) \in X$ such that the gross surpluses from $\sigma(l)$ are $(v_1(l), v_2(l))$, or else (b) specify an action $x' \in X$ and ask agent 1 to specify an integer l' such that $\sigma(l') = x'$. After agent 1 responds, agent 2 can either challenge or not. If not, the outcome is (x, y). If he challenges, he is in effect denying in case (a) that the action $\sigma(l)$ actually generates gross surpluses $(v_1(l), v_2(l))$, and in case (b) that the action l' generates gross surpluses $(v_1(l'), v_2(l'))$. In either case, the challenges are exactly as in Section 4.

Let us assume that the contract to be implemented is *strongly welfare-neutral*, in the sense that it picks the same utilities in states θ and θ' whenever the utility possibility sets are the same in those two states (in which case we will say that θ and θ' are weakly equivalent). Strong welfare-neutrality ensures that our parsimonious mechanism implements the contract.

Note that strong welfare-neutrality is indeed stronger than welfare-neutrality because two states can be weakly equivalent without being equivalent.[12] Still, Theorem 2 can be extended to strong welfare-neutrality if we replace equivalence by weak equivalence. Moreover strong welfare-neutrality is satisfied by the economic models in the literature.

8. RENEGOTIATION

One possible objection to the mechanisms constructed in Sections 4 and 5 is that they are vulnerable to renegotiation. That is, even though in equilibrium the outcomes of these mechanisms are efficient, this is not so off the equilibrium path. Moreover, the out-of-equilibrium inefficiencies are important because they deter agents from deviating from their equilibrium strategies. If agents anticipated that all inefficient outcomes would be renegotiated, the mechanisms' equilibria might collapse.

Although we agree that in practice most private contracts do not prohibit renegotiation, (and both of us have invoked the possibility of renegotiation from time to time in our own work), we should point out that the assumption that renegotiation cannot be

12. Here is an example in which the optimal contract is not strongly welfare-neutral. There is an infinite number of indescribable actions. Agent 1 is the "agent" and agent 2 the "principal." The agent can work or shirk at date 1. If he works, half the actions yield (1, 5) and the other half (0, 4). If he shirks, a third of the actions yield (2, 3) and the other two thirds yield (1, 2). So, working costs 1 to the agent and increases the principal's utility by 2. The two states, which are identified with the two efforts, are not equivalent, and so a contract that induces effort is vacuously welfare-neutral. On the other hand, the two states of nature are weakly equivalent and any strongly welfare-neutral contract induces shirking. Indeed, distinguishing between the two states requires the specification of an infinite number of actions.

prevented is motivated in the literature by considerations that lie outside the existing models.

It is not obvious why rational agents must allow renegotiation to constrain them, *i.e.* why they cannot simply write an irrevocability clause into their contracts (indeed, we see such clauses in practice in the case of irrevocable trusts). One answer sometimes offered is that under our legal system, the contract enforcer (*i.e.* the court) is not brought into the picture unless one of the contracting parties sues the other. Thus, there is no one to stop the parties if they choose to tear up their old contract and write a new one. (This point of view is formally embodied in the renegotiation models of Hart and Moore (1988, 1999), Maskin and Moore (1999), and Segal (1995)). However, parties could in principle register their contract publicly and play out the mechanism before an arbitrator.[13] Indeed, the current legal system already accommodates similar arrangements in the realm of labour negotiations.

An alternative argument offered for renegotiation is that, even if parties can successfully prevent themselves from invalidating their original contract, they may be able to write a second separate contract that "undoes" some of the provisions of the first. Of course, for such a scenario to be problematic, there would have to be more than one enforcement authority (a unique authority could simply invalidate any second contract between the parties)[14], which raises the question of why enforcement is not centralized. But even putting that issue aside, there seems a straightforward way that parties can commit themselves not to write subsequent contracts. Namely, they can write a clause into the original contract stating that if either party produces evidence of a second contract (or an equivalent set of contracts with third parties,[15] see footnote 14), he is entitled to collect a large penalty from the other party, and that in this circumstance the second contract is not enforceable (where precedence is defined by the registration date). In this way, the parties can ensure that, were they to write a second contract, they would find themselves in a "prisoner's dilemma" situation, in which each had the incentive to inform on the other.

As with our discussion of transactions costs, we do not mean to suggest that renegotiation is irrelevant in practice. Indeed, we acknowledge that it is both important and pervasive. Although we wish only to point out that a rational theory underlying renegotiation is still lacking, we remain hopeful that one (perhaps based on bounded rationality) may yet be constructed. In this respect we agree with Hart–Moore (1999), which contains an interesting discussion of the renegotiation assumption. Moreover, despite the fact that

13. One might object to such an arrangement by pointing out that the cost of hiring an arbitrator could make contracting prohibitively expensive. However, the need for arbitrators can be dispensed with as long as (i) the actions that parties take in the implementing mechanism can be verified after the fact and (ii) the original contract is registered and contains a "penalty" clause, as in the text below, ensuring that if either party produces evidence of an attempt to recontract he can collect a large fine from the other. Such a contract need not involve the courts or arbitrators unless one of the parties appeals to them after the terms of the contract have already been violated.

14. Actually, this is not necessarily true, as Oliver Hart pointed out to us; in some circumstances the parties might be able to undo the first contract through a set of contracts with third parties rather than through direct transactions. For example, suppose that *ex post* the buyer and seller of some good want to exchange more of that good than the original contract specified. Instead of the seller directly selling the buyer an additional amount, they could go through a middleman. And since the arrangements with the middleman would be *first* contracts, they would not be invalid under a "no second contract" rule.

15. Alternatively, the contract could ban *all* subsequent transactions with third parties. This would, in general, be inefficient, since some trades with third parties might be legitimate and have nothing to do with the initial contract. However, the contract could take care to ban such transactions only *off* the equilibrium path, so that the ban would serve simply as a punishment for deviation.

allowing for renegotiation may be at odds with full rationality, we remain interested in analyzing the effect of renegotiation on the irrelevance theorem.

To model renegotiation, let us adhere to Hart and Moore (1988, 1998), Maskin and Moore (1999) and Segal (1995, 1999) by assuming that:

(i) the contract is not registered publicly,

(ii) if the mechanism prescribed by the contract generates an *ex post* inefficient outcome, the parties renegotiate this outcome and move to a Pareto dominating point on the Pareto frontier,

(iii) renegotiation is "payoff-relevant." That is, the point that is reached on the Pareto frontier is entirely determined by the payoffs at the inefficient outcome and by the von Neumann–Morgenstern equivalence class of the number-based payoff functions. [As we will see, this implies that if two states are equivalent before renegotiation, they are still equivalent after renegotiation.]

There are *two types of irrelevance theorems* that can be looked for: One can search for conditions under which neither renegotiation nor indescribability constrains the set of implementable payoffs. Alternatively, one can consider cases in which renegotiation possibly restricts the set of implementable allocations, but in which indescribability does not further delimit feasible payoffs.[16]

We first obtain a result of the first type. We show that, under risk aversion and unbounded transfers, renegotiation quite generally does not constrain the set of implementable allocations even if states are indescribable. The key assumption is renegotiation welfare-neutrality, which is somewhat stronger than the ordinary welfare-neutrality invoked in Theorem 1, and, like its predecessor, guarantees that the payoffs to be implemented do not depend on information that cannot be elicited when actions are indescribable. As in Sections 4 and 5, we make further assumptions that guarantee that an agent's utility can be decreased (the agent can be punished) when he misrepresents the state of nature and is challenged. In Sections 4 and 5, this punishment could take the form of an inefficient allocation, or, relatedly, of a transfer to a third party. However, such punishments are vulnerable to renegotiation (in the case of an inefficient allocation) or to collusion between the two agents (in the case of transfers to third parties). To devise punishments that are immune to renegotiation and/or collusion, we assume agents are (at least slightly) risk averse and, as in Section 4, that penalties can be large. Thus, we make the following assumptions:

(a) for all $\theta \in \Theta$, utility functions take the form

$$u_i^\theta(x, y_i) = U_i(w_i^\theta(x) + y_i) \quad \text{for } i = 1, 2,$$

where $U_i: \mathbb{R} \to \mathbb{R}$ is increasing and strictly concave;

(b) there is no bound on the magnitude of individual agents' transfers

$$Y = \{(y_1, y_2) | y_1 + y_2 = 0\}.$$

Suppose, to simplify matters, that for each state θ, there is a unique efficient action x^θ and that the action profile (x, y_1, y_2) is renegotiated to the point $(x^\theta, y_1 + \Delta y_1^\theta(x), y_2 + \Delta y_2^\theta(x))$ where $\Delta y_1^\theta(x)$ and $\Delta y_2^\theta(x)$ $(\Delta y_1^\theta(x) + \Delta y_2^\theta(x) = 0)$ depend on x and θ, but not on

16. The latter sort of irrelevance theorem asserts that a second form of contract incompleteness, namely that due to renegotiation, does not interfere with incompleteness due to indescribability. Similar irrelevance theorems may hold when other reasons for contract incompleteness such as uncoordinated contracts (multiprincipal situations) are introduced, but we have not investigated the matter.

(y_1, y_2). (The assumption that the renegotiation is independent of (y_1, y_2) is strong but is invoked only for expositional convenience).

Renegotiation ensures that each agent i ends up with the same surplus $w_i^\theta(x^\theta)$ in state θ, regardless of the action x and the transfers (y_1, y_2) that result as the outcome of a mechanism. In keeping with requirement (ii) that renegotiation lead to Pareto improvements, we assume that

$$w_i^\theta(x^\theta) + \Delta y_i^\theta(x) \geqq w_i^\theta(x) \quad \text{for } i = 1, 2. \tag{10}$$

As for requirement (iii) that renegotiation be payoff-relevant, consider two equivalent states θ and θ'. By definition, for $i = 1, 2$, there exist $a_i > 0$, β_i and a bijection $\pi: X^\theta \to X^{\theta'}$ with $\pi(x^\theta) = x^{\theta'}$ such that

$$U_i(w_i^\theta(x) + y_i) = a_i U_i(w_i^{\theta'}(\pi(x)) + y_i) + \beta_i, \tag{11}$$

for all $x \in X^\theta$ and all y_i. Now, for any (y_1, y_2) the outcome (x, y_1, y_2) in state θ is equivalent to the outcome $(\pi(x), y_1, y_2)$ in state θ'. The requirement of payoff-relevant renegotiation demands that these two outcomes still be equivalent after renegotiation; *i.e.*,

$$U_i(w_i^\theta(x^\theta) + y_i + \Delta y_i^\theta(x)) = a_i U_i(w_i^{\theta'}(x^{\theta'}) + y_i + \Delta y_i^{\theta'}(\pi(x))) + \beta_i. \tag{12}$$

But (11) and (12) imply that

$$\Delta y_i^\theta(x) = \Delta y_i^{\theta'}(\pi(x)), \tag{13}$$

for all $x \in X^\theta$.

In the no-renegotiation case (Theorem 1) the crucial hypothesis ensuring that a complete context be implementable was welfare-neutrality. When renegotiation is possible, *renegotiation welfare-neutrality* becomes the pertinent condition. To define this, we shall call two states θ and θ' *renegotiation equivalent* if there exists a bijection $\pi: X^\theta \to X^{\theta'}$ with $\pi(x^\theta) = x^{\theta'}$ such that

$$\Delta y_i^\theta(x) = \Delta y_i^{\theta'}(\pi(x)) \quad \text{for } i = 1, 2, \quad \text{for all } x \in X^\theta.$$

Notice, from (13), that two states are renegotiation equivalent if they are equivalent in the ordinary sense. The converse, however, is not true (see below for a discussion of this).

We now define a contract f to be *renegotiation welfare-neutral* if, whenever θ is renegotiation equivalent to θ', $f(\theta) = (x^\theta, y^\theta)$ and $f(\theta') = (x^{\theta'}, y^{\theta'})$ where x^θ and $x^{\theta'}$ are the efficient actions in states θ and θ', and

$$y^\theta = y^{\theta'}.$$

Because renegotiation equivalence is weaker than ordinary equivalence, renegotiation welfare-neutrality is stronger than ordinary welfare-neutrality. Nevertheless, it is not a very restrictive assumption in typical incomplete contract models. Consider, for example, a buyer-seller framework in which in each state θ, there is one "relevant" good and a number of "irrelevant" goods that could be produced by the seller. Suppose that it is uniquely optimal for the buyer to consume one unit of the relevant good. The buyer's valuation is then $V(e_B, \kappa)$ and the seller's *ex post* production cost $C(e_S, \kappa)$ where e_B and e_S are the investments by the buyer and seller and κ is a random variable that is realized before trade. Let $\theta = (e_B, e_S, \kappa)$ be the "state of nature." Denote "no trade" by x^0 and suppose that it yields zero surplus: $w_i^\theta(x^0) = 0$ for all i and θ. Assume further that renegotiation follows a generalized Nash bargaining process: The buyer and the seller receive fractions τ_B and τ_S of the gain from renegotiating ($\tau_B + \tau_S = 1$). In particular, at the no

trade outcome total surplus is zero. Hence the total gain from renegotiating is

$$w_B^\theta(x^\theta) + w_S^\theta(x^\theta) = V(e_B, \kappa) - C(e_S, \kappa),$$

and so the buyer's and seller's payoffs after renegotiating from the no-trade point are, respectively,

$$V(e_B, \kappa) + \Delta y_B^\theta(x^0) = \tau_B[V(e_B, \kappa) - C(e_S, \kappa)] \tag{14}$$

and

$$-C(e_S, \kappa) + \Delta y_S^\theta(x^0) = \tau_S[V(e_B, \kappa) - C(e_S, \kappa)]. \tag{15}$$

Hence, (13)–(15) imply that a necessary condition for two states $\theta = (e_B, e_S, \kappa)$ and $\theta' = (e'_B, e'_S, \kappa')$ to be renegotiation equivalent is thus that

$$\tau_B C(\theta) + \tau_S V(\theta) = \tau_B C(\theta') + \tau_S V(\theta'). \tag{16}$$

Suppose that the set $Z = \{(c, v) |$ there exists θ such that $(C(\theta), V(\theta)) = (c, v)\}$ contains m elements. Then for generic choices of these m elements, (16) holds only if

$$(C(\theta), V(\theta)) = (C(\theta'), V(\theta')). \tag{17}$$

But if, as in Segal (1995) and Hart–Moore (1999), the irrelevant goods' values and costs are independent of the state, then two states satisfying (17) are equivalent in the usual sense. Hence, the fact that (16) implies (17) means that renegotiation welfare-neutrality is no stronger than ordinary welfare-neutrality. In other words, there exist first-best efficient contracts in Segal's model satisfying renegotiation welfare-neutrality.

We can now state:

Theorem 3 *Assume that utility functions take the form* $u_i^\theta(x, y_i) = U_i(w_i^\theta(x) + y_i)$ *for* $i = 1, 2$ *with* U_i *increasing and strictly concave, that there is no bound on feasible transfers:* $Y = \{(y_1, y_2) | y_1 + y_2 = 0\}$, *that there exists a describable action* x^0 *such that* $x^0 \in X^\theta$ *for all* θ, *and that the contract* f *is Pareto-optimal and renegotiation welfare neutral. Then the number-based contract* $\hat{f}$ *associated with* f *can be implemented in subgame perfect equilibrium subject to renegotiation when states are indescribable.*

Proof. See Appendix 2. ||

There are two senses in which implementing a contract is potentially harder when renegotiation is possible than when it is not. The first is that renegotiation welfare-neutrality is, in principle, stronger than ordinary welfare-neutrality. Even so, we argued above that, in typical incomplete-contract models, renegotiation welfare-neutrality is not very restrictive. We also already alluded to the second difficulty; namely, that of punishing parties for deviating from equilibrium. This unlike the first problem, is a serious constraint in many models of the literature. However these models invariably assume that parties are risk-neutral, and it turns out that when this assumption is relaxed the problem entailed in devising punishments vanishes. To see how renegotiation can make punishment problematic, recall the mechanism of Section 4. There, agent 1 pays a fine P to agent 2 if challenged by agent 2. And agent 2 pays a fine $2P$ to a third party, if his challenge turns out to be unsuccessful. If renegotiation is possible, however, then the agents will write a new contract rather than turn over any money to the third party. In this new contract, agent 1 will presumably get part of the $2P$ that would otherwise have gone to the third

party. Thus he may well be better off inducing the challenge to fail even if the challenge was valid. Hence renegotiation interferes with our mechanism in a fundamental way.

Suppose, however, that we modify the mechanism so that, if his challenge fails, agent 2 pays his fine to agent 1 rather than to a third party. Moreover, assume that the magnitude of the fine is a random variable, whose realization is determined only after the challenge fails. If the agents are both risk neutral, then the randomness of the fine is irrelevant; only its expectation matters. Moreover, the fact that agent 1 receives the fine means that he has the incentive to cause a valid challenge to fail. Suppose, however, that at least agent 2 is risk averse. Then we can find a lottery such that, for agent 1, the certainty equivalent of the payment he receives is zero, but, for agent 2, the certainty equivalent of the payment he makes is negative. Indeed, by adjusting the lottery, we can make the latter certainty equivalent as negative as we like. Hence, by penalizing agent 2 for failed challenges in this way, we can ensure that he does not have the incentive to challenge falsely, while at the same time avoiding giving agent 1 the incentive to make a valid challenge fail.

Now, one might ask, shouldn't the randomness in agent 2's penalty cause the parties to renegotiate it beforehand?[17] Notice, however, that if agent 2 has indeed made a valid challenge, then in the continuation equilibrium, the penalty will not be invoked, and so renegotiation is not an issue. Indeed, agent 2 would resist renegotiating since otherwise he might give agent 1 the incentive to cause the valid challenge to fail. On the other hand, if agent 2 has challenged falsely, then the parties *should* anticipate the penalty will be invoked. Hence, in this case, they will want to renegotiate before the challenge succeeds or fails. Nevertheless, if the penalty is severe enough, then, even after renegotiation, agent 2 will still be worse off than had he refrained from challenging falsely. So he will be deterred from doing so. Details are provided in Appendix 2.

Even though the proof of Theorem 3 makes use of large penalties (as in the example of Section 4), we conjecture that, as in the proof of Theorem 1, these are not needed, *i.e.*, that arbitrarily small penalties (combined with an arbitrarily small degree of risk aversion) will do.

More generally, though, once the possibility of renegotiation is admitted, the set of implementable payoffs when states are describable may be reduced. The pertinent issue for us, however, is whether or not this set is further reduced by indescribability. Hence we simply *assume* that a contract is implementable (under renegotiation) when the states are describable and ask whether the contract remains implementable (under renegotiation) when the states are indescribable.

We provide conditions under which indescribability makes no difference at all. Let us say that the set of states Θ is *maximal* if for all θ, $\theta' \in \Theta$ such that $|X^{\theta}| = |X^{\theta'}|$ and all bijections $\pi: X^{\theta} \to X^{\theta'}$, there exists $\theta'' \in \Theta$ such that $X^{\theta''} = X^{\theta}$ and $u^{\theta''}(x, y) = u^{\theta'}(\pi(x), y)$ for all $(x, y) \in X^{\theta} \times Y$. Note that maximality *per se* is a weak assumption since the state θ'' need not have high probability (as we will see, however, the assumptions of maximality and welfare neutrality are together quite restrictive).

Now, when renegotiation *à la* Assumptions (i)–(iii) is possible, we can introduce a "renegotiation function" $h^{\cdot}(\cdot, \cdot)$, where $h^{\theta}(x, y) \in X^{\theta} \times Y$ is the Pareto-optimal alternative to which (x, y) is renegotiated in state θ. For all θ, let $\tilde{u}_i^{\theta}(x, y) = u_i^{\theta}(h^{\theta}(x, y))$. Then $h^{\theta}(x, y)$ must satisfy

$$\tilde{u}_i^{\theta}(x, y_i) \geqq u_i^{\theta}(x, y_i), \qquad i = 1, 2.$$

17. Another possiblity (raised by Oliver Hart) is that one party or the other might attempt to obtain insurance against the lottery from a risk-neutral third party. But such an insurance arrangement could be forestalled if the parties kept the details of the randomizing device (*e.g.* a computer program) to themselves.

Moreover, the payoff-relevance requirement (iii) dictates that if $\theta \sim \theta'$—so that for some bijection $\pi: X^{\theta} \to X^{\theta'}$ and scalars $a_i > 0, \beta_i$, condition (2) holds—and

$$u_i^{\theta}(x, y_i) = a_i u_i^{\theta'}(\pi(x), y_i) + \beta_i, \qquad i = 1, 2,$$

then

$$\tilde{u}_i^{\theta}(x, y_i) = a_i \tilde{u}_i^{\theta'}(\pi(x), y_i) + \beta_i, \qquad i = 1, 2.$$

Theorem 4. *Suppose that the set of describable states is maximal. Assume that the number-based contract $\hat{f}$ corresponds to a complete contract f that is renegotiation welfare-neutral, is Pareto optimal but not one of the extreme points on the Pareto frontier, and is implementable in subgame-perfect equilibrium when states are describable and renegotiation is allowed (according to Assumptions* (i)–(iii)*). Then $\hat{f}$ can also be implemented in subgame-perfect equilibrium subject to renegotiation when states are indescribable.*

Theorem 4, which is proved in Appendix 3, indicates that the key to the indescribability being irrelevant is, once again, a form of welfare-neutrality. We now show however, that renegotiation welfare-neutrality is a more demanding condition than ordinary welfare-neutrality. Recall from Theorem 2 that state-independence of the ratios of the marginal utilities of money and the unidentifiability of effort imply welfare-neutrality. We show that in general the requirement that these two assumptions be satisfied for renegotiated utilities $\tilde{u}^{\theta}(\cdot)$ is strictly stronger than they hold for $u^{\theta}(\cdot)$.

Theorem 5. (i) *Suppose that state-independence of the ratios of the marginal utilities of money and unidentifiability of effort conditions hold under renegotiation, that is for payoffs $\tilde{u}^{\theta}(\cdot)$. Then, these two conditions also hold in the absence of renegotiation, that is for payoffs $u^{\theta}(\cdot)$.*

(ii) *However, the converse is not true: either (or both) of the two conditions may hold in the absence of renegotiation, yet be violated under renegotiation.*

Proof. (i) First recall that, from payoff-relevant renegotiation, if $\theta \sim \theta'$, then $\theta \overset{r}{\sim} \theta'$, where $\theta \overset{r}{\sim} \theta'$ means that the two states are equivalent for the payoffs functions $\tilde{u}(\cdot)$ that obtain under renegotiation.

Suppose that the state-independence condition holds under renegotiation. Then, if $\theta \overset{r}{\sim} \theta'$, there exist scalars $\tilde{\alpha} > 0$, $\tilde{\beta}_1, \tilde{\beta}_2$, and a bijection $\pi: X^{\theta} \to X^{\theta'}$ such that

$$\tilde{u}_i^{\theta}(x, y_i) = \tilde{\alpha} \tilde{u}_i^{\theta'}(\pi(x), y_i) + \tilde{\beta}_i \quad \text{for all } x \in X^{\theta}, y_i \in Y_i \text{ and } i = 1, 2.$$

Suppose, furthermore, that $\theta \sim \theta'$. Then, for $i = 1, 2$, there exist scalars $a_i > 0$ and β_i, and a bijection $\pi: X^{\theta} \to X^{\theta'}$ such that

$$u_i^{\theta}(x, y_i) = a_i u_i^{\theta'}(\pi(x), y_i) + \beta_i \quad \text{for all } x \in X^{\theta}, y_i \in Y_i \text{ and } i = 1, 2.$$

But, $u^{\theta}(\cdot)$ and $\tilde{u}^{\theta}(\cdot)$ coincide on the Pareto frontier as do $u^{\theta'}(\cdot)$ and $\tilde{u}^{\theta'}(\cdot)$. And because there are at least two points on that frontier by assumption, we conclude that $a_1 = a_2$ and so the state-independence condition also holds in the absence of renegotiation.

That the unidentifiability of effort under renegotiation implies that effort is also unidentified in the absence of renegotiation results directly from the definition of unidentifiability and from the fact that two states that are equivalent in the absence of renegotiation are also equivalent under renegotiation.

(ii) That the converse does not hold is established through an example; see below. ||

Example showing that the state-independence and the unidentified-effort assumptions are stronger under renegotiation:

There are two states of nature, θ and θ', and four actions, a_1, a_2, a_3, a_4. Let $X^{\theta} = \{a_1, a_2, a_3\}$ and $X^{\theta'} = \{a_1, a_2, a_4\}$. The action-based payoff matrices (with columns corresponding to actions and rows to agents) are

$$\begin{matrix} a_1 & a_2 & a_3 \end{matrix}$$
$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \end{bmatrix} \text{ in state } \theta, \text{ and } \begin{matrix} a_1 & a_2 & a_4 \end{matrix} \begin{bmatrix} 0 & 1 & -1 \\ 2 & 0 & -3 \end{bmatrix} \text{ in state } \theta'.$$

Note that $\theta \nsim \theta'$: the only bijection from X^{θ} to $X^{\theta'}$ that preserves the ordering of preferences is $\pi: \pi(a_1) = a_2$, $\pi(a_2) = a_1$, $\pi(a_3) = a_4$, but condition (2) is violated for any $\{a_2, \beta_2\}$. Because the two states are not equivalent, state-independence holds vacuously in the absence of renegotiation. Now consider renegotiation and assume that the ineffi cient action a_3 is renegotiated to a_1 in state θ and similarly a_4 is renegotiated to a_2 in state θ'. So, the action-based payoff matrices under renegotiation are

$$\begin{matrix} a_1 & a_2 & a_3 \end{matrix}$$
$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ in state } \theta, \text{ and } \begin{matrix} a_1 & a_2 & a_4 \end{matrix} \begin{bmatrix} 0 & 1 & 1 \\ 2 & 0 & 0 \end{bmatrix} \text{ in state } \theta'.$$

Hence for the permutation π defined above.

$$\tilde{u}_1^{\theta}(a) = \tilde{u}_1^{\theta'}(\pi(a)) \quad \text{and} \quad \tilde{u}_2^{\theta}(a) = \tfrac{1}{2}\tilde{u}_2^{\theta'}(\pi(a)).$$

So, $\theta \overset{r}{\sim} \theta'$, and state-independence fails to hold under renegotiation ($a_1 = 1 \neq a_2 = \frac{1}{2}$). It is also easy to see (indeed, the hypotheses of Theorem 1 are satisfied) that one can implement the contract that prescribes action a_1 in both states θ and θ' (an efficient "risk-sharing" arrangement) in the absence of renegotiation, whether or not actions are describable. This contract can also be implemented under renegotiation if actions are describable (the mechanism can simply specify the constant action a_1), but cannot be implemented under renegotiation in the case of indescribable actions (from welfare neutrality, agent 2 must have payoff 1 in state θ if he has payoff 2 in state θ').

This example also illustrates the point that the unidentified-effort assumption is more easily satisfied in the absence of renegotiation. Suppose that agent 1, say, chooses between efforts e and e' and that the two states θ and θ' result from e and e', respectively. Because $\theta \nsim \theta'$, the unidentified-effort assumption is vacuously satisfied in the absence of renegotiation. On the other hand, $\theta \overset{r}{\sim} \theta'$ and

$$p(\theta | e, \{\theta, \theta'\}) = 1 \neq p(\theta | e', \{\theta, \theta'\}) = 0,$$

and so the unidentified-effort assumption is not satisfied under renegotiation.

Thus Theorem 5 suggests that indescribability of actions is more likely to matter when renegotiation is allowed. Intuitively, under dynamic programming, indescribability is analogous to a "Blackwell garbling" of the information structure. This garbling is more costly when contract designers cannot use the threat of Pareto-suboptimal actions in order to elicit the state of nature.

Although Theorem 3 and 4 provide sufficient conditions under which indescribability does not matter, future research ought to delineate more carefully when it *does* matter. We must emphasize, however, that the fact that indescribability may interfere with contract efficiency does not, by itself, justify a focus on specific classes of contracts. More

work remains to be done in order to unveil the exact implications of indescribability for contract design.

9. DISCUSSION

What morals should we draw from the irrelevance theorem? One possible reaction is to dismiss the mechanisms constructed in the proof of Theorems 1 and 3 as hopelessly "unrealistic." Indeed, the lack-of-realism charge is sometimes leveled at the implementation literature more generally. In our view, however, this criticism is misguided. The implementation literature makes no pretense that anything resembling its general mechanisms are or should be used in practice. [One reason for the complexity of the mechanisms in this literature is that they are constructed to handle very general environments.] Rather it is an effort to determine which allocations are consistent with contracting by rational agents and which are not. Specifically, Theorems 1 and 3 demonstrate that in a large class of cases, the transaction costs of foreseeing contingencies have no bearing on implementable contracts. Hence, in our view, a more justifiable reaction to Theorems 1 and 3 would be to call for a weaker form of rationality. Unfortunately, our profession has, for the most part, made little progress toward modelling bounded rationality in a satisfactory way.[18]

If we are to explain "simple institutions" such as property rights, authority (or more generally, decision processes), short-term contracts, and so forth, a theory of bounded rationality is certainly an important, perhaps ultimately essential, ingredient. But for now, it is not the only reasonable approach as we argue below. In the short run there are really two options: to focus on simple institutions by assumption, or to reject the conventional wisdom that complete contract theory is incapable of explaining simple institutions.

Focus on simple institutions

One way to proceed is to decide at the outset to consider nothing but "simple contracts" or "simple institutions." Our view is that there is nothing wrong with this approach as long as we are aware of its shortcomings (insiders are quite aware of its limitations, but outsiders or new entrants in the field need not be).

The first difficulty relates to the definition of "simplicity." Is an *ex post* auction of an asset, a contract indexed on the value of publicly traded financial assets, or an option-to-sell contract complex? Are message games such as those observed in tenure reviews or regulatory hearings complex? It does not seem so; yet such apparently simple contracting clauses are rarely allowed in the literature. To be sure, the spirit of the literature is to incorporate such clauses if they seem relevant to the context, but there is no algorithm that enables us to know when we have exhausted the range of "simple contracting opportunities."

Second, and relatedly, there is the question of the value of the short-cut. All economic models are caricatures of reality. They provide a metaphor that can be used to better understand the world, to suggest reforms, and to derive testable implications. The insights we have so far gleaned from incomplete contract models (restricting attention to ownership contracts, authority relationships, short-term contracts, *etc.*) are quite encouraging. But, as usual, taking a short-cut (here restricting the set of feasible contracts on *a priori* grounds rather than from first principles) requires faith that what actually justifies the

18. For a recent attempt in the context of contracts, see MacLeod (1996).

short-cut does not interact with the rest of the reasoning. In the history of economic thought, examples abound in which taking a short-cut turned out to be a rewarding exercise (*e.g.* developing general equilibrium theory before having a model of large-number-of-traders price formation, or building the paradigms of moral hazard and adverse selection before formally introducing costs of information acquisition to explain the asymmetry of information); sometimes, however, conclusions do not seem robust to an endogenization of the short-cut (*e.g.* some of the pre-rational-expectations macroeconomics). At this stage, the literature on simple institutions seems to fall within the former category, but we must await more fundamentalist modeling in order to confirm this intuition.

How far can complete contract theory take us?

A number of recent papers have argued that in certain *very structured environments*, the optimal complete contract outcome could be implemented through an apparently incomplete contract or simple institution. Before we briefly discuss these, we should point out that the purpose of these exercises is not to show that "simple contracts" are in fact optimal; indeed the papers must invoke extreme assumptions in order to obtain such results. Rather, this literature attempts to derive circumstances in which, *under full rationality*, the loss from using simple contracts rather than optimal complex ones is small. In such circumstances, one would expect bounded rationality considerations to dominate such losses and weigh in favor of simple contracts over slightly more efficient, but also more complex contracts. So, the approach consists of identifying factors that reduce the power of complex contracts relative to simple ones, with the hope that, with the advent of a theory of bounded rationality, small amounts of bounded rationality will indeed induce contracts to be simple.

Several papers fall within this "complete contracting approach to simple institutions." We already mentioned in the introduction papers showing that in specific environments, the optimal complete contract is equivalent to a complete absence of contract or to a simple looking contract such as a property right or an authority relationship. A second branch has looked for complete contract foundations for short-term contracts. Aghion–Bolton (1987), Hermalin (1988), Diamond (1993) and others have argued in various contexts that parties who contract with private information may want to signal that they are "not afraid of returning to the market in the future" in order to obtain better terms of trade today, and that signing a short-term contract is the practical way of doing so.[19] Third, authority relationships (or more generally decision processes) may find a natural habitat in complete contracting models in which at least one party responds little to monetary incentives. Private benefits attached to decisions are then the primary determinants of the parties' incentives to participate and to behave. Put simply, the parties then care primarily about who will have his say in picking decisions, and it is therefore not surprising that authority-like contracts emerge from the analysis.[20] In contrast, strong responsiveness to monetary incentives creates scope for more efficient schemes in which messages and monetary transfers are used to take total-surplus-maximizing decisions and to compensate parties for foregone private benefits. One would then expect contracts to differ substantially from simple authority relationships. As a last illustration, one can,

19. Laffont and Tirole (1993, Ch. 16) analyze an overlapping-regulators environment and show that, in some circumstances, the optimal regulatory contract with a firm is a short-term contract despite the fact that such a contract creates ratcheting. The idea is that the corresponding inefficiency is dominated by the possibility for the new regulator to "undo" an improper contract signed by a captured former regulator.

20. See Aghion–Tirole (1997) and Tirole (1999) on this point.

following Aghion–Bolton (1992), view the financial structure of firms as packages of return streams and decision rights. One can then consider that the return stream attached to a claim is but an incentive scheme designed to induce holders of this claim to interfere in the "right way" (that is, in a way that properly disciplines the insiders), an idea well in line with standard complete contract theory.[21]

To sum up, we do not argue that complete contract theory *per se* explains the observed simplicity of contracts, but rather that it provides clues as to when bounded rationality considerations (yet to be investigated) are likely to dominate inefficiencies linked with the simplicity restriction.

APPENDIX

1. *Verification of the Moore and Repullo* (1988) *Sufficient Conditions*

Claim. *If Assumptions* 1 *and* 2 *are satisfied, then, for any action space* X, *there exists a successful mechanism* g^X, *i.e., a mechanism such that, for all pairs of number-based payoff functions* $v \in V$ (*where* v_i: $M \times Y_i \to \mathbb{R}$, $i = 1, 2$) *for which* $|M| = |X|$, *and for all deciphering keys* σ: $M \to X$, *the unique subgame-perfect equilibrium payoffs of the game* g^X *when agents have utility functions* $(v_1(\sigma^{-1}(\cdot), \cdot), v_2(\sigma^{-1}(\cdot), \cdot))$ *are* $v(\hat{f}(v))$.

Proof. We are dealing with a specified action set X and so standard implementation theory applies. We will show that the sufficient conditions for subgame-perfect implementation in Moore and Repullo (1988) are satisfied.

Let us assume that for any $v \in V$ with $v: M \times Y_i \to \mathbb{R}^2$, for any X for which $|M| = |X|$, and for any deciphering key $\sigma: M \to X$, there exists $\theta \in \Theta$ such that

$$u^\theta(\cdot, \cdot) = v(\sigma^{-1}(\cdot), \cdot).$$

This condition was not hypothesized. However, if it is violated, then there are fewer payoff functions to deal with, and so implementation is all the easier. That is, the claim holds *a fortiori*. For any X, let $\Theta_X = \{\theta \in \Theta | X^\theta = X\}$.

The first condition that we must verify is:

(a) Existence of actions for which there is "preference reversal."

We must show that, for all states $\theta, \varphi \in \Theta_X$, where u^θ is not von Neumann–Morgenstern equivalent (VNM-equivalent) to u^φ, there exist actions $a(\theta, \varphi)$, $b(\theta, \varphi) \in X \times Y$ and agents $i_1(\theta, \varphi)$ and $i_2(\theta, \varphi)$ such that

$$u^\theta_{i_1(\theta, \varphi)}(\hat{f}(v^\theta)) \geq u^\theta_{i_1(\theta, \varphi)}(a(\theta, \varphi)), \tag{A.1}$$

$$u^\theta_{i_2(\theta, \varphi)}(a(\theta, \varphi)) \geq u^\theta_{i_2(\theta, \varphi)}(b(\theta, \varphi)), \tag{A.2}$$

and

$$u^\varphi_{i_2(\theta, \varphi)}(a(\theta, \varphi)) < u^\varphi_{i_2(\theta, \varphi)}(b(\theta, \varphi)). \tag{A.3}$$

That is, we must show that there exist actions for which there is preference reversal between states θ and φ. In particular, we want agent i_2 to weakly prefer $a(\theta, \varphi)$ to $b(\theta, \varphi)$ in state θ (A.2), but to strictly prefer $b(\theta, \varphi)$ to $a(\theta, \varphi)$ in state φ (A.3). Such a preference reversal allows Moore and Repullo to build a mechanism in which each agent first announces the state of the world θ^i, and can challenge these announcements (θ^1, θ^2) by announcing φ if $\theta^1 = \theta^2$. There are four crucial properties to the construction. The first is (i) if $\theta^1 \neq \theta^2$, the outcome is a "bad" action for both agents. Next, if $\theta^1 = \theta^2 = \theta$, then (ii) provided that these announcements are true, the outcome is $\hat{f}(\theta)$ if there is no challenge and a choice between $a(\theta, \varphi)$ and $b(\theta, \varphi)$ by agent $i_2(\theta, \varphi)$ if there is a challenge by agent $i_1(\theta, \varphi)$. From (A.2) agent $i_2(\theta, \varphi)$ will choose $a(\theta, \varphi)$ and so from (A.1), agent $i_1(\theta, \varphi)$ does not gain by the challenge in state θ. But (iii) if the announcements $\theta^1 = \theta^2 = \theta$ are false, then either player can challenge with the true state φ. Moreover, from (A.3), agent $i_2(\theta, \varphi)$ now would choose $b(\theta, \varphi)$ over $a(\theta, \varphi)$, and the threat of this choice destroys the continuation equilibrium in which either $a(\theta, \varphi)$ or $b(\theta, \varphi)$ are the outcomes. In fact, the mechanism can be constructed so that the only possible continuation equilibrium is one in which $i_1(\theta, \varphi)$ chooses his favorite alternative subject to the other agent getting at least the utility he would from a "bad" outcome. Finally, (iv) the mechanism has these properties so that if the other agent anticipates that $i_1(\theta, \varphi)$ will

21. For instance, the debt-equity financial structure obtained in Dewatripont–Tirole (1994) can alternatively be interpreted as an incomplete or complete contract outcome.

challenge then he can "pre-challenge" and therefore choose his own favourite action, but if agent $i_1(\theta, \varphi)$ anticipates this, he can pre-pre-challenge etc.

Conditions (i) and (iii) ensure that we cannot have an equilibrium in which $\theta^1 \neq \theta^2$: The outcome is "bad" when the announcements differ but if, say, θ^2 is false then agent 1 can change his announcement to θ^2 and simultaneously challenge so as to get his favourite alternative. Condition (iv), however, guarantees that there cannot be an equilibrium in which both agents ananounce the same false state: each agent would try to pre-challenge the other. Finally, condition (ii) ensures that truthful relevation by both agents *is* an equilibrium (and, given the foregoing argument, the *only* equilibrium).

In view of conditions (i)–(iv) it remains to show that there exist (a) $a(\theta, \varphi)$ and $b(\theta, \varphi)$ satisfying (A.1)–(A.3), (b) a "favourite alternative" for each agent (actually, each agent may have multiple favourite alternatives, *i.e.* a *maximal set*) and (c) a "bad" action.

To demonstrate (A.1)–(A.3), we note that if u^θ is not VNM-equivalent to u^φ, there exists agent $i_2(\theta, \varphi)$, actions (x, y), (x', y'), (x'', y'') and scalar $\alpha \in [0, 1]$ such that

$$\alpha u^\theta_{i_2(\theta,\varphi)}(x, y) + (1 - \alpha) u^\theta_{i_2(\theta,\varphi)}(x'', y'') = u^\theta_{i_2(\theta,\varphi)}(x', y'), \tag{A.4}$$

and

$$\alpha u^\varphi_{i_2(\theta,\varphi)}(x, y) + (1 - \alpha) u^\varphi_{i_2(\theta,\varphi)}(x'', y'') < u^\varphi_{i_2(\theta,\varphi)}(x', y'). \tag{A.5}$$

Let $\hat{a}(\theta, \varphi)$ be a randomization between (x, y) and (x'', y'') with probabilities α and $1 - \alpha$ respectively. Let $\hat{b}(\theta, \varphi) = (x', y')$. Then (A.4) and (A.5) can be rewritten as

$$u^\theta_{i_2(\theta,\varphi)}(\hat{a}(\theta, \varphi)) = u^\theta_{i_2(\theta,\varphi)}(\hat{b}(\theta, \varphi)) \tag{A.6}$$

and

$$u^\varphi_{i_2(\theta,\varphi)}(\hat{a}(\theta, \varphi)) < u^\varphi_{i_2(\theta,\varphi)}(\hat{b}(\theta, \varphi)). \tag{A.7}$$

Now, for $\beta \in (0, 1)$, let $a(\theta, \varphi)$ be a randomization between (x^0, y^0) (with probability β) and $\hat{a}(\theta, \varphi)$ (with probability $1 - \beta$). Let $b(\theta, \varphi)$ be a randomization between (x^0, y^0) (with probability β) and $\hat{b}(\theta, \varphi)$ (with probability $1 - \beta$). Notice that as long as $\beta < 1$, (A.6) and (A.7) continue to hold when $\hat{a}(\theta, \varphi)$ and $\hat{b}(\theta, \varphi)$ are replaced by $a(\theta, \varphi)$ and $b(\theta, \varphi)$. Hence (A.2) and (A.3) hold. Moreover for β near enough 1, (A.5) implies that

$$u_i^{\theta'}(a(\theta, \varphi)) > \tfrac{1}{2} u_i^{\theta'}(x^0, y^{00}) + \tfrac{1}{2} u_i^{\theta'}(x^0, y^0) \quad \text{for all } \theta', i, \tag{A.8}$$

and

$$u_i^{\theta'}(b(\theta, \varphi)) > \tfrac{1}{2} u_i^{\theta'}(x^0, y^{00}) + \tfrac{1}{2} u_i^{\theta'}(x^0, y^0) \quad \text{for all } \theta', i. \tag{A.9}$$

Now, for β sufficiently close to 1, there exists $i = i_1(\theta, \varphi)$ such that

$$u^\theta_{i_1(\theta,\varphi)}(f(\theta)) \geqq u^\theta_{i_1(\theta,\varphi)}(a(\theta, \varphi)), \tag{A.10}$$

as (A.1) claimed.

(b) Existence of maximal set.

Let

$$Q = \bigcup_{\theta, \varphi \in \Theta_X} \{f(\theta), a(\theta, \varphi), b(\theta, \varphi)\},$$

with the provision that if a mechanism ever specifies an outcome in Q, either agent has the alternative option of implementing a 50–50 randomization between (x^0, y^0) and (x^0, y^{00}) instead. That is, in any state $\theta \in \Theta_X$, an alternative $a \in Q$ is really a contingent alternative $\hat{a}$ in which

$$\hat{a} = \begin{cases} a, \text{ if } u_i^\theta(a) \geqq \tfrac{1}{2} u_i^\theta(x^0, y^0) + \tfrac{1}{2} u_i^\theta(x^0, y^{00}), & \text{for } i = 1, 2, \\ \tfrac{1}{2}(x^0, y^0) + \tfrac{1}{2}(x^0, y^{00}), & \text{otherwise.} \end{cases} \tag{A.11}$$

The next condition to verify is that there exists a subset $B \subseteq X \times Y$ containing Q such that, for $i = 1, 2$ and all $\theta \in \Theta_X$, if

$$M_i(\theta) = \{a \in B \mid u_i^\theta(a) \geqq u_i^\theta(b) \quad \text{for all } b \in B\},$$

then

$$M_i(\theta) \text{ is non-empty,} \tag{A.12}$$

$$M_i(\theta) \cap M_j(\theta) \text{ is empty,} \tag{A.13}$$

and

$$M_i(\theta) \cap Q \text{ is empty.} \tag{A.14}$$

For state θ, the sets $M_1(\theta)$ and $M_2(\theta)$ are, respectively, 1's and 2's favourite alternatives in the set B, which includes the optimal alternative $f(\theta)$ as well as the preference reversal alternatives. $a(\theta, \varphi)$ and $b(\theta, \varphi)$ for each φ.

To demonstrate (A.12)–(A.14), let $R = X \times Y - Q$ with the provision that if a mechanism ever specifies an alternative in R, each agent has the option of implementing a 25–75 randomization between (x^0, y^0) and (x^0, y^{00}) instead. That is, in any state $\theta \in \Theta_X$, an alternative $a \in R$ is really a contingent alternative $\hat{a}$ in which

$$\hat{a} = \begin{cases} a, \text{ if } u_i^\theta(a) \geqq \frac{1}{4}u_i^\theta(x^0, y^0) + \frac{3}{4}u_i^\theta(x^0, y^{00}), & i = 1, 2 \\ \frac{1}{4}(x^0, y^0) + \frac{3}{4}(x^0, y^{00}), & \text{otherwise.} \end{cases} \tag{A.15}$$

Take

$$B = Q \cup R.$$

As long as X and Θ_X are finite, the continuity of $u_i^\theta(x, y)$ ensures that (A.12) holds for all $\theta \in \Theta_X$ and $i = 1, 2$. Now, if $a \in M_1(\theta)$, then, from Assumption 1 (ii) and Assumption 2,

$$u_1^\theta(a) > v_1^\theta(\hat{f}(v^\theta)),$$

and so, from the Pareto-optimality of $\hat{f}$, $v_2^\theta(\hat{f}(v^\theta)) > u_2^\theta(a)$. Hence, $a \notin M_2(\theta)$, and we conclude that (A.13) holds.

Finally, to establish (A.14), consider $a \in Q$. Then, for some θ', $\varphi' \in \Theta$,

$$a = \begin{matrix} f(\theta') \\ \text{or} \\ a(\theta', \varphi') \\ \text{or} \\ b(\theta', \varphi'). \end{matrix} \tag{A.16}$$

If θ is the true state of the world, then as we have noted, a mechanism designating the outcome a is really prescribing outcome $\hat{a}$ satisfying (A.11). From (A.11) and (A.16) we see that there are four cases:

(i) $\hat{a} = f(\theta')$;
(ii) $\hat{a} = a(\theta', \varphi')$;
(iii) $\hat{a} = b(\theta', \varphi')$;

and

(iv) $\hat{a} = \frac{1}{2}(x^0, y^0) + \frac{1}{2}(x^0, y^{00})$.

Consider case (i) first: $\hat{a} = f(\theta')$. Then, from (A.11),

$$u_j^\theta(f(\theta')) \geqq \tfrac{1}{2}u_j^\theta(x^0, y^0) + \tfrac{1}{2}u_j^\theta(x^0, y^{00}), \quad j = 1, 2. \tag{A.17}$$

From Assumption 2, the definition of R, and (A.17), there exists for each $i = 1, 2$, $a^i \in R$ such that

$$u_i^\theta(a^i) > u_i^\theta(f(\theta')), \tag{A.18}$$

and

$$u_j^\theta(a^i) > \tfrac{1}{4}u_j^\theta(x^0, y^0) + \tfrac{3}{4}u_j^\theta(x^0, y^{00}), \quad j \neq i, \tag{A.19}$$

and $f(\theta') \notin M_i(\theta)$. Next consider case (ii): $\hat{a} = a(\theta', \varphi')$. Then, (A.11) implies that

$$u_j^\theta(a(\theta', \varphi')) \geq \tfrac{1}{2}u_j^\theta(x^0, y^0) + \tfrac{1}{2}u_j^\theta(x^0, y^{00}), \quad j = 1, 2. \tag{A,.20}$$

From Assumption 1 (iii) and (A.20) there exists, for each $i = 1, 2$, $a^i \in R$ satisfying $u_i^\theta(a^i) > u_i^\theta(a(\theta', \varphi'))$ and (A.19). Hence $a(\theta', \varphi') \notin M_i(\theta)$. Similarly, in case (iii), $b(\theta', \varphi') \notin M_i(\theta)$.

Finally, consider case (iv): $\hat{a} = \frac{1}{2}(x^0, y^0) + \frac{1}{2}(x^0, y^{00})$. In that case $v_i^\theta(\hat{f}(\theta)) > u_i^\theta(\hat{a})$ for $i = 1, 2$, and so, once again, $\hat{a} \notin M_i(\theta)$ for $i = 1, 2$. We conclude that (A.14) holds, as claimed.

(c) Existence of "bad" actions.

To conclude the verification of the Moore–Repullo conditions, we must show that there exists an action a^* such that for all $\theta \in \Theta_x$ and $i = 1, 2$,

$$u_i^\theta(a) > u_i^\theta(a^*) \quad \text{for all } a \in M_1(\theta) \cup M_2(\theta) \cup Q. \tag{A.21}$$

But from the above choices of $M_i(\theta)$ and Q, it is easy to see that (A.21) is satisfied when $a^* = (x^0, y^{00})$. ||

2. *Proof of Theorem* 3

Consider the following mechanism:
(i) agent 1 announces

- an action set X
- payoff functions $w_i: X \to \mathbb{R}$, $i = 1, 2$,
- an outcome $(x^{\hat\theta}, y^{\hat\theta})$ where $(x^{\hat\theta}, y^{\hat\theta}) = f(\hat\theta)$ and $\hat\theta = (X, w_1, w_2)$.

[Note that the assumption that f is welfare neutral and renegotiation ensure that it is possible to determine what $f(\hat\theta)$ is without $\hat\theta$ having been described in advance.]

Choose P big enough so that

$$\Delta y_1^\theta(x) - P < y_1^\theta$$

and

$$\Delta y_2^\theta(x) + P > y_2^\theta$$

for all θ and $x \in X^\theta$. [Note that here we are again invoking the assumption that f is welfare neutral.]

If agent 1 fails to do the above, the outcome is $(x^0, -P, P)$.
(ii) agent 2 can challenge or not

- if not, the outcome is $(x^{\hat\theta}, y^{\hat\theta})$
- if he does challenge, agent 1 must pay him P
- he can either challenge X (in which case we proceed as usual) or (w_1, w_2)
- if he challenges w_1, then he must choose $x, x' \in X$ and $y \in \mathbb{R}$ such that

$$\Delta y_1^{\hat\theta}(x) < y_1^{\hat\theta}(x') + \gamma.$$

Agent 1 can chooose between x and x'. If he chooses x, the challenge succeeds (agent 2 proves that agent 1's claim that the state is $\hat\theta$ is false) and the outcome is $(x, 0, 0)$ (before renegotiation). If he chooses x', the challenge fails; and agent 1 should end up with utility (assuming $\hat\theta$ is the true state)

$$w_1^{\hat\theta}(x^{\hat\theta}) + \Delta y_1^{\hat\theta}(x') + \gamma - P.$$

We want to punish agent 2 for having challenged falsely. Accordingly choose $K > 0 > L$ so that

$$\tfrac{1}{2} U_1(w_1^{\hat\theta}(x^{\hat\theta}) + \Delta y_1^{\hat\theta}(x') + K - P) + \tfrac{1}{2} U_1(w_1^{\hat\theta}(x^{\hat\theta}) + \Delta y_1^{\hat\theta}(x') + L - P) = U_1(w_1^{\hat\theta}(x^{\hat\theta}) + \Delta y_1^{\hat\theta}(x') + \gamma - P).$$

For $|L|$ very big, K will also have to be very big. Indeed note that given our assumption about strict risk aversion, for a sequence satisfying the previous equation

$$\lim_{|L| \to \infty} \{\tfrac{1}{2} U_2(w_2^{\hat\theta}(x^{\hat\theta}) + \Delta y_2^{\hat\theta}(x') - K + P) + \tfrac{1}{2} U_2(w_2^{\hat\theta}(x^{\hat\theta}) + \Delta y_2^{\hat\theta}(x') - L + P)\} = -\infty.$$

Accordingly, at the same time that agent 2 challenges, have him choose K and L (satisfying the equality above) so big that

$$\tfrac{1}{2} U_2(w_2^{\hat\theta}(x^{\hat\theta}) + \Delta y_2^{\hat\theta}(x') - K + P) + \tfrac{1}{2} U_2(w_2^{\hat\theta}(x^{\hat\theta}) + \Delta y_2^{\hat\theta}(x') - L + P) < U_2(w_2^{\hat\theta}(x^{\hat\theta}) + y_2^{\hat\theta}).$$

Then if agent 2's challenge fails, the outcome is $(x^{\hat\theta}, K - P + \Delta y_1^{\hat\theta}(x'), -K + P + \Delta y_2^{\hat\theta}(x'))$ with probability 1/2 and $(x^{\hat\theta}, L - P + \Delta y_1^{\hat\theta}(x'), -L + P + \Delta y_2^{\hat\theta}(x'))$ with probability 1/2. Hence agent 2 is deterred from challenging falsely. If agent 1 chooses x', the outcome of the lottery is known immediately, so there is no renegotiation once agent 1 has chosen x'. Thus agent 1 cannot choose x' in order to "blackmail" agent 2 and force renegotiation to something he prefers to the x choice. [Because the (K, L) lottery entails risk, it is inefficient and, therefore, one might expect it to be possibly renegotiated before agent 1's choice between x and x'. Note, however, that if agent 2 has challenged validly, the (K, L) lottery won't be invoked, and so there is no need to renegotiate it. Indeed, agent 2 will refuse to renegotiate it to eliminate any incentive agent 1 might have to falsely demonstrate that agent 2's challenge is invalid. If agent 2's challenge is invalid, then the (K, L) lottery will be renegotiated, but as long as $|K|$ and $|L|$ are big enough, agent 2 will still be worse off than if he had not challenged.]

A symmetric argument applies if agent 2 challenges w_2. ||

3. *Indescribability and Renegotiation*

Notice that, for a Pareto-optimal complete contract f as defined in Section 2, renegotiation is not an issue: $f(\theta)$ is Pareto for each state θ, and so there is nothing to renegotiate. Let us therefore drop the assumption that utilities are verifiable in the describable states case (*i.e.* let us now suppose that states are describable but not verifiable). Hence, as in Section 4, agents have to resort to a mechanism in order to implement a complete

contract. However, since agents can describe the possible action sets, the mechanism can be made contingent on the realized action space. Specifically, we shall denote a contingent mechanism by $\{g^X\}_X$ where, for each possible X,

$$g^X: S_1^X \times S_2^X \to X \times Y,$$

and S_i^X is agent i's strategy space when $X^\theta = X$. We shall say that $\{g^X\}_X$ *implements* the complete contract f in the describable states case if, for all $\theta \in \Theta$ and all subgame-perfect equilibria (s_1^θ, s_2^θ) in state θ,

$$u_i^\theta(g^{X^\theta}(s_1^\theta, s_2^\theta)) = u_i^\theta(f(\theta)), \quad i = 1, 2.$$

We first establish a general lemma and then apply it to the case of renegotiation.

Lemma 1. *Suppose that the set of states Θ is maximal. If the number-based contract $\hat{f}$ corresponds to a complete contract f that is welfare-neutral, implementable in subgame-perfect equilibrium in the describable states case, is Pareto optimal and not one of the two extreme points on the Pareto frontier, then $\hat{f}$ can be implemented in subgame-perfect equilibrium when states are indescribable.*

Proof. Suppose that f is welfare-neutral and is implemented by $\{g^X\}_X$ in the case where actions are describable in advance. When states are not describable, agents can no longer describe the elements of the set X in advance. However, they can regard these as *dummy* sets (in the same way that actions are dummy variables in number-based contracts).

Accordingly, consider the following three-stage mechanism. In the first stage agent 1 announces (i) a set of actions $\hat{X}$, (ii) a set of dummy actions X with $|X| = |\hat{X}|$, and (iii) a bijection $\gamma: X \to \hat{X}$. In stage 2, agent 2 can accept or challenge. If he accepts, the mechanism moves to stage 3. If he challenges, then he must either exhibit a feasible $x \notin \hat{X}$ or an infeasible $x \in \hat{X}$, in which case he has the right to implement any feasible action of his choosing and the game ends. If the game goes to stage 3, the agents play the mecanism g^X "transformed by γ" (the γ-transformed g^X): If they play strategies (s_1, s_2) such that

$$g^X(s_1, s_2) = (x, y), \tag{A.22}$$

then the outcome is

$$(\gamma(x), y). \tag{A.23}$$

If at any point some agent i fails to follow these rules, then the other agent gets the opportunity to choose any feasible alternative he wishes.

We claim that this mechanism implements $\hat{f}$. To see this, suppose first that agent 1 has announced the true action space $\hat{X}$ as well as dummy action space X and bijection $\gamma: X \to \hat{X}$, and that agent 2 has accepted. Let $\hat{\theta} \in \Theta$ be the true state of nature.

Now perform the thought-experiment in which, instead of a dummy set, agent 1 announces a real set X. Then, from maximality, there exists $\theta \in \Theta$ such that $X^\theta = X$ and

$$u^\theta(x, y) = u^{\hat{\theta}}(\gamma(x), y). \tag{A.24}$$

Now, because, by hypothesis, g^X implements f when the action space is X,

$$u^\theta(g^X(s^*)) = v^\theta(\hat{f}(\theta)), \tag{A.25}$$

for any equilibrium s^* in state θ. But from (A.24), s^* is an equilibrium for g^X in state θ if and only if it is a continuation equilibrium for the γ-transformed g^X in state $\hat{\theta}$. Hence, from (A.24) and (A.25) the equilibrium payoffs of the γ-transformed g^X in state $\hat{\theta}$ are $v^\theta(\hat{f}(\theta))$. But from welfare-neutrality and (A.24)

$$v^{\hat{\theta}}(\hat{f}(\hat{\theta})) = v^\theta(f(\theta)). \tag{A.26}$$

Hence, regardless of the γ announced, the continuation equilibrium payoffs of our mechanism are $v^{\hat{\theta}}(\hat{f}(\theta))$, and so agent 1 can announce any such bijection. We have been arguing as though X were a real action set. But note that everything continues to go through if X consists only of numbers.

It remains only to note that if agent 1 fails to announce the true action space $\hat{X}$, then agent 2 can successfully challenge him, and the assumption that $f(\theta)$ is on the Pareto frontier, but not one of the extreme points, implies that he does worse than (A.26). Hence (A.26) constitutes the overall equilibrium payoffs, that is, the mechanism implements $\hat{f}$.

Turning to the case of renegotiation, and given renegotiation function $h^{\cdot}(\cdot, \cdot)$ described in the text, mechanism g and state $\theta \in \Theta$, let

$$g^{\theta}(s_1, s_2) = h^{\theta}(g(s_1, s_2))$$

for all feasible strategies (s_1, s_2). It is straightforward to verify that each step of the proof of Lemma 1 continues to hold when we replace each mechanism g with g^{θ}, and replace welfare-neutrality with renegotiation welfare-neutrality. In particular, note that payoff-relevant renegotiation implies that maximality holds for the utility functions $\tilde{u}_i^{\theta}$ provided that it does for the utility functions u_i^{θ}.

Acknowledgements. The authors are grateful to Oliver Hart, Jean-Jacques Laffont, John Moore, Jean-Charles Rochet and a referee for helpful discussions and comments.

REFERENCES

AGHION, P. and BOLTON, P. (1992). "An Incomplete Contracts Approach to Financial Contracting", *Review of Economic Studies*, **59**, 473–493.

AGHION, P. and TIROLE, J. (1997), "Formal and Real Authority in Organizations", *Journal of Political Economy*, **105**, 1–29.

AGHION, P., DEWATRIPONT, M. and REY, P. (1994), "Renegotiation Design with Unverifiable Information", *Econometrica*, **62**, 257–282.

ARROW, K. (1953), "Le Rôle des Valeurs Boursières pour la Répartition la Meilleure des Risques", *Econométrie* (Paris: Centre National de la Recherche Scientifique) [Translated as: Arrow, K. (1964), "The role of Securities in the Optimal Allocation of Risk Bearing", *Review of Economic Studies*, **31**, 91–96].

CHE, Y. K. and HAUSCH, D. (1998), "Cooperative Investments and the Value of Contracting: Coase vs Williamson", *American Economic Review* (forthcoming).

DEWATRIPONT, M. (1989), "Renegotiation and Information Relevation over Time: The Case of Optimal Labor Contracts", *Quarterly Journal of Economics*, **104**, 589–620.

DEWATRIPONT, M. and TIROLE, J. (1994), "A Theory of Debt and Equity: Diversity of Securities and Manager-Shareholder Congruence", *Quarterly Journal of Economics*, **109**, 1027–1054.

DIAMOND, D. (1993), "Seniority and Maturity of Debt Contracts", *Journal of Financial Economics*, **33**, 341–368.

FUDENBERG, D. and TIROLE, J. (1990), "Moral Hazard and Renegotiation in Agency Contracts", *Econometrica*, **58**, 1279–1320.

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

HART, O. (1995) *Firms, Contracts, and Financial Structure*, (Oxford: Oxford University Press).

HART, O. and MOORE, J. (1990), "Property Rights and the Nature of the Firm", *Journal of Political Economy*, **98**, 1119–1158.

HART, O. and MOORE, J. (1999), "Foundations of Incomplete Contracts", *Review of Economic Studies*, **66**, 115–138.

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

HERMALIN, B., (1988), "Adverse Selection, Contract Length, and the Provision of On-the-Job Training," in *Three Essays on the Theory of Contracts*, PhD thesis, Massachusetts Institute of Technology.

HOLMSTRÖM, B. and TIROLE, J. (1991), "Transfer Pricing and Organizational Form", *Journal of Law, Economics and Organization*, **7**, 201–228.

LAFFONT, J.-J. and TIROLE, J. (1993) *A Theory of Incentives in Procurement and Regulation* (Cambridge: MIT Press).

MACLEOD, B. (1996) "Complexity, Contracts, and the Employment Relationship" (Mimeo).

MASKIN, E. and MOORE, J. (1999), "Implementation and Renegotiation", *Review of Economic Studies*, **66**, 39–56.

MASKIN, E. and TIROLE, J. (1999), "Two Remarks on the Property Rights Literature", *Review of Economic Studies*, **66**, 139–149.

MOORE, J. (1992), "Implementation in Environments with Complete Information", in *Advances in Economic Theory*, J. J. Laffont, (ed.), (Cambridge: Cambridge University Press) p. 182–282.

MOORE, J., and REPULLO, R. (1998), "Subgame Perfect Implementation", *Econometrica*, **56**, 1191–1220.

NÖLDEKE, G. and SCHMIDT, K. (1995), "Option Contracts and Renegotiation: A Solution to the Hold-Up Problem", *Rand Journal of Economics*, **26**, 163–179.

NÖLDEKE, G. and SCHMIDT, K. (1998), "Sequential Investments and Options to Own", (Mimeo, University of Bonn and Munich).

PALFREY, T. (1998), "Implementation Theory", in R. Aumann and S. Hart (eds.) *Handbook of Game Theory*, vol. 3 (Amsterdam: North Holland).

SEGAL, I. (1995) *Essays on Commitment, Renegotiation, and Incompleteness of Contracts*, PhD thesis, Harvard University.

SEGAL, I. (1999), "Complexity and Renegotiation: A Foundation for Incomplete Contracts", *Review of Economic Studies*, **66**, 57–82.
TIROLE, J. (1999), "Incomplete Contracts: Where Do We Stand?" *Econometrica* (forthcoming).
WERNERFELT, B. (1989), "Unforeseen Contingencies and Market Failure", (Mimeo, MIT).
WILLIAMSON, O. (1975) *Markets and Hierarchies: Analysis of Antitrust Implications* (New York: Free Press).
WILLIAMSON, O. (1985) *The Economic Institutions of Capitalism* (New York: Free Press).