Limited Observability as a Constrint in Contract Design

Stefan Krasa
University of Illinois at Urbana−Champaign

Steven R. Williams
University of Illinois at Urbana−Champaign

Abstract

Limited observability is the assumption that economic agents can only observe a finite amount of information. Given this constraint, contracts among agents are necessarily finite and incomplete in comparison to the ideal complete contract that we model as infinite in detail. We consider the extent to which finite contracts can approximate the idealized complete contracts. The objectives of the paper are: (i) to identify properties of agents' preferences that determine whether or not finiteness of contracts causes significant inefficiency: (ii) to evaluate the performance of finite contracts against the ideal optimal contracts in principal−agent and bilateral bargaining models.

Published: 2001
URL: http://www.business.uiuc.edu/Working_Papers/papers/01−0128.pdf

Limited Observability as a Constraint in Contract Design*

Stefan Krasa and Steven R. Williams
Department of Economics
University of Illinois
1206 South Sixth Street
Champaign IL 61820
http://www.staff.uiuc.edu/~skrasa
swillia3@uiuc.edu

October 12, 2001

Abstract

Limited observability is the assumption that economic agents can only observe a finite amount of information. Given this constraint, contracts among agents are necessarily finite and incomplete in comparison to the ideal complete contract that we model as infinite in detail. We consider the extent to which finite contracts can approximate the idealized complete contracts. The objectives of the paper are: (i) to identify properties of agents' preferences that determine whether or not finiteness of contracts causes significant inefficiency; (ii) to evaluate the performance of finite contracts against the ideal optimal contracts in principal-agent and bilateral bargaining models.

Keywords*:* limited observability, bounded rationality, contractual incompleteness

JEL Classification Numbers: D82, C78

1 Introduction

People are limited in their ability to observe and to process information. Contracts among individuals and institutions are therefore limited in length and

*This paper originated in In-Koo Cho's seminar on bounded rationality that we attended in the fall of 1998. We thank In-Koo for inviting us to participate in his seminar and for his comments concerning this paper. We thank David S. Williams for helpful conversations concerning key person insurance and Nabil Al-Najjar, Jim Peck, and Stan Reiter for their comments and suggestions.

may fail to address all of the different contingencies that may arise. This incompleteness can lead to inefficiency in the contractual outcome, as evidenced by legal disputes or costly renegotiation. We develop in this paper a model that encompasses both the limited contracts that are used in practice and the ideal contracts that address all contingencies. The goal of this paper is to identify properties of agents' preferences that determine whether or not limits on contractual length cause significant inefficiency in comparison to an ideal contract.

To illustrate the issues that we address, consider the central administration (or center) and a division manager of a firm. The issue to be contracted is the production plan of the division. A contingency is any random variable whose value affects the output of the firm. The ability of each worker ("high" or "low") in the division is a contingency. For simplicity, we assume in this example that the abilities of a countable number of workers constitute all contingencies. The division manager can only observe the abilities of a finite number of workers, which limits the center and the division manager to finite contracts. Together with the countability of the set of contingencies, this models the idea that the center and the division manager cannot address all contingencies in their contract. The choice of a production plan is a principal-agent problem with hidden information in which the center is the principal and the division is the agent. An optimal production plan in this problem typically depends upon an infinite number of the workers' abilities.

It is quite possible that beyond a certain number of workers the total impact upon output of the abilities of all other workers is relatively small. We state this formally as a continuity condition on the output function: the total effect of all but a finite number n of contingencies can be made arbitrarily small by making n sufficiently large. Not surprisingly, continuity of this kind is useful for proving that an infinite contract can be approximated arbitrarily closely with finite contracts. The optimal production plan in this example might thus be approximated with a finite contract that does not address an infinite number of workers whose abilities are insignificant in their cumulative effect upon output.

Such a contract, however, might still be quite lengthy if the inefficiency due to contractual finiteness is to be made acceptably small. Contracts between a center and a division, however, may be much simpler. A contract may address only the abilities of a few key employees. The fact that the abilities of only some number n of employees are deemed sufficiently important in their effect upon output does not explain the use of contracts that are nearly state independent, such as delegation of decision making to the division manager. Our answer lies in incentive compatibility and, more precisely, in properties of the output function that determine whether or not an incentive constraint is binding. If the production plan is to address a particular worker's ability, then the division manager must be induced to reveal what he observes concerning that worker. Roughly, a particular worker's ability is *reversible* if the abilities of other workers may be realized in a number of different ways so as to undo (or reverse) the effect upon output of the given worker's ability, regardless of its realization. Conversely, the worker's ability is *strongly irreversible* if its effect upon output cannot be masked by the abilities of others. While each of these two properties concern the relationship between a worker's ability and other contingencies that affect output, neither is simply an issue of the absolute magnitude of the worker's effect upon output. Reversibility and strong irreversibility are thus distinct from the continuity condition discussed above. A worker's ability that is reversible presents an opportunity for misrepresentation to the division manager that may be unsolvable. We show that it may not be possible for an incentive compatible and finite contract to address that worker's ability, even if its impact upon output is large. Conversely, if a worker's ability is strongly irreversible, then the division manager can be compensated so as to induce him to reveal the realized ability of the worker. A worker whose ability is strongly irreversible can therefore be addressed in an incentive compatible and finite contract.

Output is the measure of welfare in the above example. We refer to an agent's measure of well-being as *preferences* in our general model, though it may have a more precise interpretation (e.g., as output) in examples such as the one above. We thus characterize two properties (reversibility and strong irreversibility) that concern how a contingency affects an agent's preferences in relation to other contingencies. Strong irreversibility or reversibility can determine whether an incentive compatible finite contract either can or can not address the given contingency. Assuming a limited ability to observe, our paper thus grounds in agents' preferences a theory of (i) which contingencies may and which contingencies may not be addressed in an incentive compatible and finite contract, and (ii) when a substantial welfare loss necessarily occurs because of contractual finiteness.

Model and Results. Our general contracting model is developed in section 2. There are two agents. For an agent j who has access to private information and for any $n \in \mathbb{N}$, *contingency* n is a random variable whose value is either 0 or 1. This can be interpreted as a "no" or "yes" answer to question n concerning the state of the world. The *type* α_j of an agent j who has private information is a sequence $\alpha_j = (a_{j,n})_{n \in \mathbb{N}}$ in which $a_{j,n}$ is a realization of the nth contingency. The type set A_j of agent j is either a singleton set or $\{0,1\}^{\mathbb{N}}$, depending upon whether he has access to private information. The *set of states of the world* is $A = A_1 \times A_2$. The utility $u_j(\alpha, c)$ of each agent is a function of the state α and the choice c. A *contract* is a function $f\colon A \to C$ from A into a set C of possible collective choices for the agents. A contract is *finite* if it only depends on a finite number of contingencies. Otherwise, it is *infinite*.

We deviate from Harsanyi's (1967-68) model of a game of incomplete information by assuming that agent j does not fully observe his type once it is realized. *Limited observability* is the assumption that agent j can choose to costlessly observe the realization of any finite number of the contingencies that define his type but not the entire type itself when it is infinite in length. His type can thus be regarded as information that he can *access* to the extent of his bounded ability. Except for this assumption, the agents are otherwise perfectly rational in the sense that they maximize their expected return conditional upon their finite observations.[1] Limited observability is formalized in section 3.[2]

A contract between agents who are constrained by limited observability is necessarily finite. A finite contract in our paper is feasible and an infinite contract is an ideal. This is analogous to using an infinite number of traders to model perfect competition. Markets in reality never have an infinite number of traders and we are not proposing that the relevant contingencies in an actual contracting problem are truly infinite. Rather, in the same way in which an infinity of traders avoids the quantification of how rapidly market power diminishes as market size increases, an infinity of contingencies models the unattainable complexity of addressing every contingency without the need for assumptions about the feasible length of a contract or the marginal costs of lengthening it.[3]

In the same sense that a perfectly competitive market is meaningful as an abstraction to the extent that it can be approximated by finite markets, an infinite contract is meaningful only if it is approached in the limit by a sequence of finite contracts. Our interest here is primarily in whether or not there are costs associated with contractual finiteness. A contract is *recordable* if its efficiency is matched or surpassed by the limiting performance of some sequence of finite contracts. We have selected the term "recordable" to suggest that the gist of the infinite contract can be written in a finite contract, with only details omitted, to whatever degree of accuracy is sought. If a contract is not recordable, then finite contracts are bounded away from the contract in the welfare measure of the problem. Recordability is a rather weak requirement to impose on an infinite contract as a way of modeling bounds on contractual length. It is far less severe than the alternative of simply restricting attention to finite contracts.

We investigate recordability in sections 5 and 6 in two common contracting problems with incomplete information. The first problem is a principal-agent problem with hidden information and the second is a generalization of the Chatterjee-Samuelson (1983) bilateral bargaining problem. Reflecting the

[1] The term "limited observability" is drawn from Radner (2000), which classifies the various approaches to bounded rationality in economic theory. Our approach is closest in spirit to his category of "costly rationality models" in that an agent in our model is rational in his acquisition and use of a finite number of bits of information, given that (i) there is no cost to observing any finite number of contingencies, and (ii) it is infinitely costly to fully observe his type. We thus follow the Savage approach of modeling bounded rationality as rationality subject to the costs of collecting and processing information. This cost function for observing contingencies, however, is not explicitly studied within the paper.

[2] While it may not capture all aspects of a human being's capacity for observation, our model of the relationship between an agent and the world is motivated by interpreting the agent as an empirical researcher. The selection of a finite number of contingencies to observe is the design of an experiment, and the realization of those contingencies is a data set. Data must necessarily be expressable as a sequence of *bits* (or binary digits) for scientific analysis (e.g., using statistics or a computer). Our assumption that an agent's type is a sequence of bits thus simply assumes that data about the state of the world is collected in the reduced form in which it must ultimately be presentable. This interpretation was suggested to us by Nabil Al-Najjar.

[3] As emphasized in Radner (2000), determining the cost of acquiring information is a difficult empirical problem whose solution may depend greatly upon the context.

presence of incomplete information, we consider the recordability of those contracts that are optimal subject to the constraints of incentive compatibility and interim individual rationality. The issue is whether or not such an optimal contract is recordable with the additional requirement that the finite contracts in the sequence converging to the optimal contract must also be incentive compatible and interim individual rational. Recordability is shown to depend crucially upon reversibility of contingencies. Similar results hold in each of the two problems:

1. The optimal contract is infinite.

2. If all contingencies are reversible, then a recordable contract necessarily achieves nothing relative to the measure of performance that determines the optimal mechanism. The optimal contract in this case is thus not recordable.

3. Conversely, if all contingencies are strongly irreversible, then the optimal contract is recordable.

Contractual incompleteness describes a situation in which a meaningful welfare loss occurs because a contract fails to address some contingencies. Result 2 is the most provocative because it describes a case in which contractual incompleteness arises endogenously in a contracting problem because of limited observability and properties of the agents' preferences. We clarify this point in the context of the bilateral bargaining problem. A choice $c \in C$ consists of a decision of whether or not the trade of a good or service takes place between a seller and a buyer along with a price at which the transaction may occur. The set A of states of the world consists of all attributes of the good that are utility relevant and about which the agents may wish to contract (e.g., time and date of delivery, aspects of warranty, and physical characteristics of the good). As characterized by a result of Myerson and Satterthwaite (1983), the optimal contract in this problem is infinite. Theorem 9 below states that if all contingencies observed by each trader are reversible, then the ex ante gains from trade equal zero in any recordable contract. Ex ante contracting is therefore completely pointless. A practical interpretation of this result is that ex ante contracting may be replaced in such a problem with interim negotiation of the terms of trade.

Finally, a sequence of examples are worked throughout the paper that concern the case in which agent j's type $\alpha_j = (a_{j,t})_{t \in \mathbb{N}}$ affects his utility through a real value $v_j(\alpha_j)$ given by the formula

$$v_j(\alpha_j) = \sum_{t=1}^{\infty} a_{j,t} \delta^t \tag{1}$$

for some $\delta \in (0,1)$. All of the ideas of this paper are illustrated with this simple family of examples by varying the common ratio δ. These examples are instructive because they illustrate some rather difficult ideas using little more than the formula for the sum a geometric series together with some arithmetic. We find it encouraging that our ideas arise in such a common family of formulas and do not require odd examples for the sake of illustration. None of the theorems in the paper, however, depend upon this special form of payoff function, and the issues that we raise are clearly not restricted to this particular family of examples.

Related Work. Our paper originates most directly in the work of Anderlini and Felli (1994, 1998), who grounded the theory of contractual incompleteness in the theory of computational complexity. The most fundamental idea that we draw from their paper is finiteness as a characteristic of real contracts and infiniteness as an ideal of contracting. We go further than Anderlini and Felli (1994) in three respects. First, limited observability deepens the modeling of the contracting process by explicitly modeling the limitations on the agents' abilities that cause contractual finiteness. Second, incomplete information and the issue of incentive compatibility are of central interest in our paper, which they did not address. Third, we start with a sequence $\alpha_j = (a_{j,t})_{t\in\mathbb{N}}$ of bits as agent j's private information, while they take the more standard approach of assuming that his information is a private value $v_j \in \mathbb{R}$. They assume the functional form of (1) for $\delta = 0.5$ to "unravel" v_j into the sequence α_j of bits that define v_j's binary expansion and then consider the computability of different values of v_j. Starting with $A_j = \{0,1\}^{\mathbb{N}}$, we instead consider all mappings $v_j : A_j \to \mathbb{R}$ that determine how aspects of the state of world affect the agent's well-being. Our approach allows us to identify properties of the agents' preferences over states of the world that distinguish contracting problems in which finite contracts can be approximately optimal from those in which they cannot, which is not possible in the Anderlini-Felli approach in which these preferences are fixed.

Similar to our approach, Segal (1999) models contracting within a complex environment by agents who are limited in their ability to describe the world. A buyer and a seller in his model can trade one of n objects, only k of which can be described. It is shown that ex ante contracting diminishes in value as n goes to infinity and the environment becomes increasingly complex, which resembles some of our results on bilateral trade. The main differences between our paper and his are (i) our focus upon private information, which he does not address, and (ii) his emphasis upon renegotiation, which we do not address. The most significant similarity between the two papers is their common reliance upon a complex environment as a key ingredient in a model of boundedly rational behavior. Segal (1999, p. 74) explains this issue as follows:

> While much has been said about the role of bounded rationality in explaining contractual incompleteness, existing models have not been able to explain how people could be irrational enough not to be able to describe all of the possible contingencies *ex ante*, yet rational enough to foresee their payoffs *ex ante* and to describe any given contingency *ex post* (see e.g. Maskin and Tirole 1999). In our view, any attempt to model bounded rationality in a simple

> environment is doomed to fall into the trap of describing decision makers as either "completely dumb" or "perfectly rational". Neither is an attractive alternative for modeling "transaction costs". It is only in environments reflecting the real world's complexities that an intermediate region of "bounded rationality" emerges.

We believe that our use of $\{0,1\}^{\mathbb{N}}$ as the set of states of the world together with the constraint of limited observability creates a model of human behavior that successfully lands within Segal's intermediate region.[4]

2 The Model

We start with a probability space $(A, \mathcal{A}, \pi)$, where A is the set of states of the world, $\mathcal{A}$ is the σ-algebra of measurable sets, and π is the common prior of the two agents. The set A_j is agent j's *type space,* with $\alpha_j \in A_j$ denoting a generic element. We shall consider both $A_i = \{\alpha_i\}$ and $A_j = \{0,1\}^{\mathbb{N}}$ (one-sided incomplete information) and $A_1 = A_2 = \{0,1\}^{\mathbb{N}}$ (two-sided incomplete information). In each case, $A = A_1 \times A_2$. Let π_j denote the marginal distribution on A_j. A type α_j of agent j is written as $\alpha_j = (a_{j,t})_{t\in\mathbb{N}}$. For $n \in \mathbb{N}$, the *initial string* $\alpha_{j,n-}$ and the *tail* $\alpha_{j,n+}$ determined by α_j and n are

$$\alpha_{j,n-} = (a_{j,t})_{1\leq t<n} \ \textit{and} \ \alpha_{j,n+} = (a_{j,t})_{t>n} \,,$$

respectively. Notice that the initial string and the tail determined by α_j and n omit the realization $a_{j,n}$ of the nth contingency. Let $A_{j,n-}$ denote the set of all initial strings of length $n-1$ and $A_{j,n+}$ the set of all tails from the $(n+1)$st contingency to infinity. All cylinder sets of the form $\{a_{j,n-}\} \times A_{j,n-1+}$ are assumed to be measurable with respect to π_j so that probabilities are well-defined conditional on the realization of any initial string $a_{j,n-}$.[5]

Let C be the choice set. A contract is a mapping $f : A \to C$.

Definition 1 *A contract f is* ***finite*** *if and only if there exists $n_1, n_2 \in \mathbb{N}$ such that $f(\alpha) = f(\alpha')$ for all $\alpha = (a_{1,t}, a_{2,t})_{t\in\mathbb{N}}$ and $\alpha' = (a'_{1,t}, a'_{2,t})_{t\in\mathbb{N}}$ with $\alpha_{i,n_i+1-} = \alpha'_{i,n_i+1-}$ for $i = 1, 2$. Otherwise, f is* ***infinite****.*

For n_1 and n_2 as described in this definition, the finite contract $f(\alpha)$ is sometimes written as $f(\alpha_{1,n_1-}, \alpha_{2,n_2-})$, reflecting the fact that $(\alpha_{1,n_1-}, \alpha_{2,n_2-})$ de-

[4] A different tact to complexifying the state space can be found in Al-Najjar (2000). While his set of states is a subinterval of the real line (as in Anderlini and Felli (1994)), Al-Najjar allows measures on this interval that are only finitely additive. This permits functions that are not computable in the sense that they cannot be approximated by simple functions.

[5] In results that concern only a single agent's private information, we will at times lessen notation by ignoring $A_i = \{\alpha_i\}$, dropping the subscript i or j denoting the individual, and writing simply $A = \{0,1\}^{\mathbb{N}}$ with generic element $\alpha = (\alpha_t)_{t\in\mathbb{N}}$. observed by one of the agents. The general formulation above is useful for our general results because it allows the cases of two-sided and one-sided incomplete information to be discussed together using the same notation.

termines $f(\alpha)$.[6]

Our discussion is limited to an *independent private value model* in the sense that:

1. The types of the agents are independent.

2. Each agent j's utility is *quasilinear* in the sense that

$$u_j(\alpha_j, c) = h_j(c)v_j(\alpha_j) + t_j(c). \tag{2}$$

The function $v_j(\alpha_j)$ is agent j's *valuation function*, which is the part of his utility that is determined directly by his type. The function $t_j(\cdot)$ is a monetary transfer to agent j and $h_j(\cdot)$ is a level or portion of $v_j(\alpha)$ that agent j receives as a result of the choice c. It is assumed throughout the paper that $h_j(\cdot)$ is bounded. The form of utility in (2) is commonly assumed in the mechanism design and the contracting literatures.

A contract f is *incentive compatible* if and only if

$$E_{A_i}\left[u_j(\alpha_j, f(\alpha))\right] \geq E_{A_i}\left[u_j(\alpha_j, f(\alpha_j^*, \alpha_i))\right] \tag{3}$$

for $j = 1, 2$ and all $\alpha_j, \alpha_j^* \in A_j$. Define

$$\begin{aligned} H_j(\alpha_j^*) &= E\left[h_j(f(\alpha)) \left| \alpha_j = \alpha_j^*\right.\right], \textit{ and} && (4) \\ T_j(\alpha_j^*) &= E\left[t_j(f(\alpha)) \left| \alpha_j = \alpha_j^*\right.\right]. && (5) \end{aligned}$$

Independence of types insures that $H_j(\alpha_j^*)$ and $T_j(\alpha_j^*)$ depend only upon the reported type α_j^* of agent j and not upon his observed type α_j. A contract f is thus incentive compatible if

$$H_j(\alpha_j^*)v_j(\alpha_j^*) + T_j(\alpha_j^*) \geq H_j(\alpha_j)\, v_j(\alpha_j^*) + T_j(\alpha_j) \tag{IC}$$

for $j = 1, 2$ and all $\alpha_j^*, \alpha_j \in A_j$. Let r_j denote agent j's reservation utility. The contract f is *interim individually rational* for agent j if

$$H_j(\alpha_j)\, v_j(\alpha_j) + T_j(\alpha_j) \geq r_j \tag{IIR}$$

for all $\alpha_j \in A_j$.

We assume for the remainder of the paper that

$$v_j(A_j) \subset \left[\underline{v}_j, \overline{v}_j\right] \quad \textit{for } i = 1, 2. \tag{6}$$

Finally, let μ_j denote the induced probability distribution on $\left[\underline{v}_j, \overline{v}_j\right]$ defined by π_j and $v_j(\cdot)$, and let $v : A \to [\underline{v}_1, \overline{v}_1] \times [\underline{v}_2, \overline{v}_2]$ denote the *valuation mapping*

$$v(\alpha) = (v_1(\alpha_1), v_2(\alpha_2)). \tag{7}$$

[6]The set of initial strings of length zero is empty, and so statements in the paper concerning $\alpha_{j,n-}$ for $n = 1$ hold vacuously. In particular, $n_i = 1$ in Definition 1 means that $f(\alpha)$ does not depend upon agent i's type α_i, and $n_1 = n_2 = 1$ means that $f(\alpha)$ is constant on A.

2.1 Recordability

Recordability indicates whether or not an infinite contract overstates the potential of contracting in the sense that its performance cannot be approximated by a sequence of finite contracts. A precise definition is given in the context of each problem considered in the remaining sections. Intuitively, a contract f is *recordable* if there exists a sequence of finite contracts $(f_t)_{t\in\mathbb{N}}$ such that

$$\lim_{t\to\infty} \|f_t\| \geq \|f\|, \tag{8}$$

where "$\|\cdot\|$" denotes a performance measure that is appropriate in the particular contracting problem. Each contract f_t in the sequence is also required to have properties appropriate to the problem. In the principal-agent problem of section 5, for instance, "$\|f\|$" is the ex ante expected payoff to the principal determined by the contract f that he offers to the agent, while in the bilateral trade problem of section 6 it is the ex ante gains from trade achieved by the traders in the contract f. The agent and the traders must willingly participate and fulfill the contract. Each f_t must therefore be both interim individually rational and incentive compatible in these problems. If f is not recordable, then all finite contracts are bounded below f according to the performance measure for the problem, i.e.,

$$\|f^*\| < \|f\| - k,$$

for some constant $k > 0$ and any finite contract f^*. A contract f that is not recordable therefore overstates the performance potential of contracting. If f is an optimal contract, then *contractual incompleteness* necessarily occurs if f is not recordable.[7]

Recordability is based upon approximating a contract f with a sequence of finite contracts that in the limit performs at least as well as f according to the performance measure $\|\cdot\|$. One might desire that the sequence $(f_t)_{t\in\mathbb{N}}$ converges in some stronger sense, e.g., that for large n the choice $f_t(\alpha)$ approximates $f(\alpha)$ in every state α. Theorems 6 and 8 below, which establish recordability of an optimal contract f, are each proven by constructing a sequence of finite contracts $(f_t)_{t\in\mathbb{N}}$ that converges pointwise to f. Our proofs of recordability of optimal contracts thus construct finite contracts f_t that not only match the performance of f but also approximately implement the same choices.

2.2 Continuity

Agent j's valuation function $v_j(\cdot)$ is *continuous* if all contingencies in a tail $\alpha_{j,n-1+}$ beyond the initial string $\alpha_{j,n-}$ that he observes are *details* in the sense that together they have only a minor effect on the value of $v_j(\cdot)$. This is formalized as follows.

[7] As stated in the Introduction, we take "contractural incompleteness" to mean simply that losses of nonnegligible magnitude occur because certain contingencies are not addressed. Anderlini and Felli (1994, p. 1098-1102) contains a deeper effort to formalize this common term.

Definition 2 *Suppose that* $A_j = \{0,1\}^{\mathbb{N}}$*. Agent* j*'s valuation function* $v_j(\cdot)$ *is* ***continuous*** *if and only if for every* $\varepsilon > 0$*, there exists an* $n \in \mathbb{N}$ *such that*

$$\left|v_j(\alpha_j) - v_j(\alpha'_j)\right| < \varepsilon \tag{9}$$

for all $\alpha_j,\ \alpha'_j \in A$ *with* $\alpha_{j,n+1-} = \alpha'_{j,n+1-}$*.*

Because of our assumption that $h_j(C)$ is bounded, continuity of the valuation function $v_j(\cdot)$ is a special case of the following condition on a general utility function $u_j(\cdot)$: for every $\varepsilon > 0$, there exists an $n \in \mathbb{N}$ such that

$$\left|u_j(\alpha_j, c) - u_j(\alpha'_j, c)\right| < \varepsilon \tag{10}$$

for all $c \in C$ and all $\alpha_j,\ \alpha'_j \in A$ with $\alpha_{j,n+1-} = \alpha'_{j,n+1-}$. This alternative formulation of continuity will prove useful later in the paper.

Continuity of $v_j(\cdot)$ as defined above is equivalent to the standard topological definition of continuity relative to the product topology on $A_j = \{0,1\}^{\mathbb{N}}$ when each set $\{0,1\}$ is assigned the discrete topology.[8] We define continuity as above both because this definition models the idea of details, which has a clear and practical meaning in the context of contracts, but also to avoid motivating this particular choice of topology on A_j.[9] In models of complete information, it can in fact be sufficient in itself (albeit given the right specification of the model) to insure recordability of optimal contracts.[10]

Example 1 *Consider* $v_j(\alpha_j) = \sum_{t=1}^{\infty} a_{j,t}\delta^t$ *for* $\delta \in (0,1)$*. If* $\alpha_{j,n-} = \alpha'_{j,n-}$*, then*

$$\left|v_j(\alpha_j) - v_j(\alpha'_j)\right| = \left|\sum_{t=n}^{\infty}\left(a_{j,t} - a'_{j,t}\right)\delta^t\right| \leq \sum_{t=n}^{\infty}\delta^t = \frac{\delta^n}{1-\delta}, \tag{11}$$

from which it is clear that $v_j(\alpha_j)$ *is continuous.*

[8] That is, $\{0\}$ and $\{1\}$ are open in $\{0,1\}$. This is true because all sets of the form $\{\alpha_{j,n-}\} \times A_{j,n-1+}$ for $n \in \mathbb{N}$ and $\alpha_{j,n-} \in A_{j,n-}$ define a base for this topology. Kelley (1955) is a classic reference for all of the topological concepts and results that are used in this paper.

[9] Continuity of $v_j(\cdot)$ also resembles *continuity at infinity of payoffs in an infinitely repeated game* when each $\alpha_{j,n-}$ can be identified with a history in the game of a particular length. This is well-known concept in repeated game theory (see, for instance, Fudenberg and Tirole (1991, Def. 4.1, p. 110)).

[10] Theorem 1 of Krasa and Williams (2000) is a result of this kind. Anderlini and Felli (1998) presents continuity conditions that are sufficient for approximating an optimal principal-agent contract with a computable contract together with examples that illustrate how various kinds of discontinuities can prevent such an approximation. These conditions are distinct from our definition of continuity because they do not concern the function v_j. Anderlini and Felli instead address other features of the principal-agent problem, including the continuity of the optimal contract, the production technology, and the principal's preferences over the agent's actions. In sum, they examine in detail the regularity conditions on the principal-agent model that we take for granted in Theorem 1 of Krasa and Williams (2000) and in section 5 below.

2.3 Contracts and Mechanisms

A goal of this paper is to investigate how preferences determine whether or not contractual incompleteness necessarily occurs in the sense that an optimal contract is infinite but not recordable. Optimization has not been previously studied for contracts in which an agent's private information is an element of $\{0,1\}^{\mathbb{N}}$. We address this issue in this subsection by connecting the theory of contracts as developed in this paper to the rich literature on optimal mechanisms. The main conclusion here is that standard results in mechanism design characterize the optimal (infinite) contracts in the principal-agent and bilateral trade models that we will consider.

A contract is a function $f : A \to C$ that determines a choice $f(\alpha)$ for each state α. Consistent with the terminology of mechanism design, a *mechanism* is a function $\hat{f} : v(A) \to C$ that determines a choice $f(v_1, v_2)$ for each pair of valuations $(v_1, v_2) \in v(A)$.[11] A mechanism $\hat{f}$ defines a contract f through composition with the valuation mapping $v(\cdot)$; a contract f defines a mechanism $\hat{f}$, however, only if it selects the same choice for all states that determine the same pair of valuations. The set of contracts is larger than the set of mechanisms, and an optimal contract may thus in principle surpass the performance of an optimal mechanism. Our objective in this subsection is to state conditions under which this does not happen, so that the optimal mechanism derived by standard methods characterizes the performance of the optimal contract.

Let

$$\Phi : C \times [\underline{v}_1, \overline{v}_1] \times [\underline{v}_2, \overline{v}_2] \to \mathbb{R}$$

be the *objective* of the contracting problem. In a principal-agent model, for instance, Φ is the principal's ex post payoff, and in a bilateral trading problem Φ is the ex post gains from trade. The *optimal contract problem* is

$$\max_{f} E_A \left[\Phi\left(f\left(\alpha\right), v(\alpha)\right)\right] \quad \textit{s.t. IC and IIR,}$$

and the *optimal mechanism problem* is

$$\max_{\hat{f}} E_{\left[\underline{v}_1, \overline{v}_1\right] \times \left[\underline{v}_2, \overline{v}_2\right]} \left[\Phi\left(\hat{f}\left(v\right), v\right)\right] \quad \textit{s.t. IC and IIR,}$$

where *IC* and *IIR* should be interpreted appropriately in the case of the optimal mechanism problem.

Theorem 11 states that if C is a convex subset of $\mathbb{R}^m$ and $\Phi(c, v_1, v_2)$ is concave in c for each v_1 and v_2, then an optimal mechanism $\hat{f}$ defines an optimal contract f through composition with the valuation mapping $v(\cdot)$.[12] Conversely,

[11] In the broader literature, "contracts" and "mechanisms" are not distinguished as they are here by their domains; the words are used almost interchangeably, depending upon the subject of the model. The distinction we make here is purely for our expositional purposes.

[12] It is also required in this theorem that h_j and t_j are affine in c for each agent j that has private information.

given these conditions on C and Φ, an optimal contract f defines an optimal mechanism $\hat{f}$ through the formula $\hat{f} \circ v = g$, where the contract $g(\alpha)$ is defined by averaging $f(\alpha^*)$ over all states α^* for which $v(\alpha^*) = v(\alpha)$.[13] A formal statement and proof of Theorem 11 is in the Appendix. The principal-agent and bilateral trade models considered in this paper satisfy the hypotheses of Theorem 11, which provides us with the information we need concerning optimal contracts in these models.

We conclude this discussion of contracts and mechanisms by addressing a fundamental question: Why are contracts written in our model on the contingencies instead of directly on the payoffs to the agents? Ultimately, the payoffs are all that matter to welfare, and hence there seems little value in addressing the contingencies that are merely the ingredients of the payoffs. Alternatively, this question can be posed as, Why not combine all types that define the same valuation into a single type?[14,15] Theorem 11 in fact provides sufficient conditions under which there is no gain in welfare from writing the contract on the underlying contingencies instead of the valuations. There are two parts of an answer to these questions. First, in many problems the choice c to be made depends not only upon the payoffs to the agents but also upon the contingencies themselves.[16] The dependence of the choice c upon the state α, however, is not the issue in the principal-agent and bilateral trade models considered in this paper, for the choice c can be selected purely on the basis of the valuations in

[13] This converse reflects the common result of mechanism design that there are no gains in ex ante performance from introducing lotteries into the operation of a mechanism. The lottery in this case is the dependence of the choice $f(\alpha^*)$ upon α^* given that $v(\alpha^*) = v(\alpha)$. Given the assumptions of the theorem, ex ante expected performance can only improve by replacing each lottery over choices $\{f(\alpha^*) \,|v(\alpha^*) = v(\alpha)\,\}$ with its certainty equivalent, which corresponds to a mechanism.

[14] This is a crucial question for our paper because it will be shown that multiple types defining the same valuation can create problems of incentive compatibility that are solvable only at the expense of introducing contractural incompleteness. This is the problem of reversible contingencies, which is defined in section 4 and then applied in sections 5 and 6. Multiple types defining the same valuation is thus a fundamental ingredient of the problem that we identify in this paper.

[15] This question has a parallel in multidimensional mechanism design, which concerns problems in which an agent's type is a vector instead of a single real number: if an agent's payoff is ultimately a single real number that is determined by the choice and the type, then why should it matter that his type is more complex than his payoff? This parallel between our approach to contracts and multidimensional mechanism design is more than coincidental, for our approach models an agent's type as countably infinite in dimension (though "dimension" is used loosely because of the discreteness of each contingency). As in multidimensional mechanism design, the increase in dimension of an agent's type correspondingly increases the number of incentive constraints. This will be become apparent in the following section on reversibility, a condition that determines when these constraints are binding.

[16] For instance, a multitude of issues may require resolution in a real bargaining problem, not merely the division of a surplus between the two agents (as is common in models of bargaining). Resolution of a labor dispute, for example, requires more than a determination of the profit of the firm and the total compensation of the labor force; there are many issues of safety, working conditions, and details of compensation that the firm and the union must resolve. While the division of the surplus may be all that matters in a welfare analysis of bargaining, determining this division may not solve the bargaining problem to the satisfaction and the needs of the bargainers.

these models. The second part of the answer concerns limited observability. An agent in our model does not know his valuation because he does not know his type. As illustrated in case 3 of Example 2 in the next section, he may not be able to place the most elementary of bounds upon it. Depending upon the formula that determines his valuation, an agent may simply be incapable of participating in a mechanism that requires him to estimate his valuation with a specified degree of accuracy. Contracts based upon exact or even approximate valuations may in this sense be unfeasible.

3 Limited Observability

The purpose of this section is to justify our focus in the remainder of the paper upon incentive compatible finite contracts. This is accomplished by grounding these contracts in an assumption concerning the limited abilities of the agents. In the case of $A_j = \{0,1\}^{\mathbb{N}}$, *limited observability by agent j of his type* is the assumption that agent j can choose to costlessly observe any initial string $\alpha_{j,n-}$ determined by his type α_j of arbitrary but finite length. There is no a priori bound on the number $n-1$ of contingencies that he may observe, and he may choose to observe different numbers of contingencies for different realizations of his type. The practical implication of this constraint is that any action or report that the agent takes conditional upon his type α_j must be determined by $\alpha_{j,n-}$ for some sufficiently large n.[17]

Consider a game in which M_j is agent j's action set and $\eta : M_1 \times M_2 \to C$ is the outcome mapping. The strategy $\gamma_j : A_j \to M_j$ of agent j is *finite* if there exists $n \in \mathbb{N}$ such that

$$\textit{if } \alpha_{j,n-} = \alpha^*_{j,n-}, \textit{ then } \gamma_j(\alpha_j) = \gamma_j(\alpha^*_j) \tag{12}$$

for all $\alpha_j, \alpha^*_j \in A_j$. The first $n-1$ contingencies of α_j thus determines $\gamma_j(\alpha_j)$ for every $\alpha_j \in A_j$. We write $\gamma_j(\alpha_j) = \gamma_j(\alpha_{j,n-})$ in this case.

An agent who is constrained by limited observability is capable of using any finite strategy. The following theorem goes further by proving that finite strategies are the only strategies that such an agent can use.

Theorem 1 *An agent who is constrained by limited observability can use a strategy if and only if the strategy is finite.*

The proofs of Theorem 1 and all other theorems are in the Appendix. The proof demonstrates that the constraint of using only finite initial strings to select actions implies the existence of a uniform length $n-1$ of initial string that is sufficient for selecting the action $\gamma_j(\alpha_j)$ for all α_j. Theorem 1 thus shows that

[17] In the case of $|A_j| = 1$ in which agent j does not have private information, both agents know agent j's unique type α_j because they know the structure of the model. Limited observability thus does not bind as a constraint on an agent who does not have private information. We write $\alpha_j = (a_{j,t})_{t \in \mathbb{N}}$ purely for notational convenience in this case; none of our results depend, however, upon a sequence representation of agent j's type in this case.

limited observability is more severe as an ex ante constraint than as an interim constraint in the following sense: while an agent can base his choice of an action on as large of a finite number of contingencies of his realized type as he wishes, a well-defined strategy ex ante defines a single upper bound on the number of contingencies that affect the value of this strategy. Theorem 1 thus reveals the limitations on an agent's actions at the interim that are imposed by the requirement of coherently specifying those actions ex ante in a strategy.

If the action $\gamma_j(\alpha_j)$ is interpreted a signal of the agent's type α_j, then a finite strategy uses only a finite set of signals. Theorem 1 thus proves that limited observability implies *limited communication* in our model, i.e., an agent's language is finite in the sense that he uses only a finite number of messages.[18] The agent also conveys at most the finite number of contingencies given in the initial string $\alpha_{j,n-}$ that determines the value of $\gamma_j(\alpha_j)$. As Example 2 at the end of this subsection reveals, limited observability is a more restrictive constraint on an agent's abilities than limited communication. The theorem also implies that a contract implemented through a game is necessarily finite if the agents are constrained by limited observability.

Limited observability also influences how we define an equilibrium in a game.

Definition 3 *A pair of strategies (γ_1, γ_2) in the game $(M_1 \times M_2, \eta)$ is a* ***Bayesian-Nash equilibrium with limited observability*** *if:*

1. *each agent's strategy is finite in the sense of (12);*
2. *for $j = 1, 2$ and assuming that γ_j depends only upon the first $n-1$ contingencies in α_j, the inequality*

$$E_A\left[u_j\left(\alpha_j, \eta\left(\gamma_j\left(\alpha^*_{j,n-}\right), \gamma_i\left(\alpha_i\right)\right)\right) \middle| \alpha_{j,t-} = \alpha^*_{j,t-}\right] \geq E_A\left[u_j\left(\alpha_j, \eta(m_j, \gamma_i\left(\alpha_i\right))\right) \middle| \alpha_{j,t-} = \alpha^*_{j,t-}\right] \quad (13)$$

holds for every $\alpha^*_j \in A_j$, $t \geq n$ and $m_j \in M_j$.

Condition 2 requires that $\gamma_j(\alpha^*_{j,n-})$ is a best response to $\gamma_i(\cdot)$ conditional upon agent j knowing *any* initial string $\alpha^*_{j,t-}$ of length $t \geq n$ that agrees with $\alpha^*_{j,n-}$ on its first $n-1$ contingencies. There are two equilibrium conditions in (13), the first reflecting agent j's ability to freely choose his action in the game and the second reflecting his ability to condition that choice upon the observation of as many contingencies as he wishes. The first is the standard condition that agent j cannot profit by deviating from $\gamma_j(\alpha^*_{j,n-})$ if observes at least enough information $\alpha^*_{j,t-}$ as may be required to compute $\gamma_j(\alpha^*_{j,n-})$. The second is that agent j does not profit from observing additional contingencies $(a^*_{j,k})_{n \leq k < t}$ beyond $\alpha^*_{j,n-}$, for doing so never provides him with grounds for profitably deviating from $\gamma_j(\alpha^*_{j,n-})$. The second condition means that an equilibrium of this kind is not simply a Bayesian-Nash equilibrium for the case in

[18] This behavioral constraint has been considered by Dow (1991), Meyer (1991), and Rubinstein ((1993),(1998, Chapter 5)).

which each agent observes only $n-1$ contingencies.[19] Notice also that unlike the standard definition of a Bayesian-Nash equilibrium, the expected values in (13) are computed not just with respect to the unknown value of the other agent's type α_i but also with respect the unknown tail $\alpha_{j,t-1+}$ of contingencies that are not observed by agent j.

Restricting attention to incentive compatible finite contracts is grounded in the constraint of limited observability through the following theorem.

Theorem 2 *Suppose that each agent's valuation function is continuous. A contract f is implemented by a Bayesian-Nash equilibrium with limited observability in some game if and only if f is finite and incentive compatible in the classical sense of (3).*

Given continuity, incentive compatible finite contracts are thus exactly those that result when the agents are constrained by limited observability. This theorem is a *revelation principle for limited observability* in the sense that the game $(M_1 \times M_2, \eta)$ that is constructed to implement a given incentive compatible finite contract f has the following properties: (i) for each agent j, $M_j = A_{j,n-}$ for some $n \in \mathbb{N}$; (ii) honest revelation of the initial string $\alpha_{j,n-}$ by each agent defines a Bayesian-Nash equilibrium with limited observability.

Limited observability also alters the constraint of interim individual rationality, though in a way that is easily seen to be inconsequential when each $v_j(\cdot)$ is continuous. Suppose the finite contract f depends only upon the initial string $\alpha_{j,n-}$ of length $n-1$ observed by each agent. The contract f is *interim individually rational given limited observability* for agent j if

$$E_A\left[u_j\left(\alpha_j, f\left(\alpha^*_{j,n-}, \alpha_{i,n-}\right)\right) \middle| \alpha_{j,t-} = \alpha^*_{j,t-}\right] \geq r_j \tag{14}$$

for all $t \geq n$ and $\alpha^*_j \in A_j$. In words, agent j's expected return conditional upon observing $\alpha^*_{j,t-}$ is at least his reservation utility r_j if he reports $\alpha^*_{j,n-}$, regardless of how many additional contingencies $(a^*_{j,k})_{n \leq k < t}$ he may choose to observe beyond the $(n-1)$st. A contract that is interim individually rational clearly satisfies (14); conversely, if $v_j(\cdot)$ is continuous, then (14) implies interim individual rationality. Imposing interim individual rationality on finite contracts is therefore consistent with limited observability.

We conclude this subsection with an example that emphasizes the impact of limited observability as a constraint on an agent's knowledge of his valuation. The reader should pay particular attention to cases 2 and 3 in this example, which show that an agent may not be able to place the most elementary of bounds upon his valuation if he is constrained by limited observability.

[19] Case 2 of Example 8 illustrates how a Bayesian-Nash equilibrium may not be a Bayesian-Nash equilibrium with limited observability precisely because of this second equilibrium condition. This example is deferred until section 6 concerning bilateral trade because it concerns the Chatterjee-Samuelson (1983) model of bargaining. A reader who is familiar with this model may wish to skip ahead to this example in order to better understand this new solution concept.

Example 2 *An agent's valuation function* $v(\alpha)$ *is*

$$v(\alpha) = \frac{1-\delta}{\delta} \sum_{t=1}^{\infty} \delta^t a_t, \tag{15}$$

where δ *will be selected below. The agent is constrained by limited observability. The formula for the sum of a geometric series along with some elementary analysis implies* $v(A) = [0, 1]$ *for* $\delta \in [0.5, 1)$.[20] *Reflecting limited communication, suppose that the agent wishes only to announce whether his valuation is high (h) in the case of* $v(\alpha) \geq 0.5$ *or low (l) in the case of* $v(\alpha) \leq 0.5$. *The formula*

$$\frac{1-\delta}{\delta} \sum_{t=2}^{\infty} \delta^t a_t \leq \frac{1-\delta}{\delta} \sum_{t=2}^{\infty} \delta^t = \frac{1-\delta}{\delta} \cdot \frac{\delta^2}{1-\delta} = \delta \tag{16}$$

is helpful in this discussion.

Case 1: $\delta = 0.5$. *It is clear from (16) that* $v(a) \geq 0.5$ *if* $a_1 = 1$ *and* $v(\alpha) \leq 0.5$ *if* $a_1 = 0$. *The strategy*

$$\gamma(\alpha) = \begin{cases} h \text{ if } a_1 = 1 \\ l \text{ if } a_1 = 0 \end{cases}$$

thus accurately communicates whether the valuation is high or low. This strategy is finite and is therefore compatible with limited observability.

Case 2*:* $\delta > 0.5$. *Formula (16) applies to show that*

$$\begin{aligned} v\left(\{a_1 = 0\} \times A_{1+}\right) &= [0, \delta], \text{ and} \\ v\left(\{a_1 = 1\} \times A_{1+}\right) &= [1-\delta, 1]. \end{aligned}$$

Both of these intervals contain 0.5 *in their interiors, and so the agent cannot identify his valuation as high or low with certainty based upon* a_1.

Case 3. *The problem observed in case 2 for* $\delta > 0.5$ *is also true of longer initial strings and alternatives to the above definitions of h and l as the language. For any* $k \geq 2$ *and any* $x \in (0, 1)$, *the agent who observes* α_{k+1-} *cannot determine with certainty for every choice of* $\alpha_{k+1-} \in A_{k+1-}$ *whether* $v(\alpha) \leq x$ *or* $v(\alpha) \geq x$. *Formally, this means that for any* $x \in (0, 1)$ *there exists an* $\alpha^*_{k+1-} \in A_{k+1-}$ *such that*

$$x \in Int\left(v\left(\left\{\alpha^*_{k+1-}\right\} \times A_{k+}\right)\right),$$

where $v(\{\alpha^*_{k+1-}\} \times A_{k+})$ *is a closed interval. The proof is in the Appendix.*

[20] The case of $\delta < 0.5$ is not considered in this example because $v(A) \subsetneq [0, 1]$ in this case, which complicates the discussion. This case is also omitted from Examples 4, 6 and 8 of sections 5 and 6. The results we invoke in these sections that characterize optimal mechanisms require regularity properties on the distribution of valuations that are not satisfied in the case of $\delta \in (0, 0.5)$.

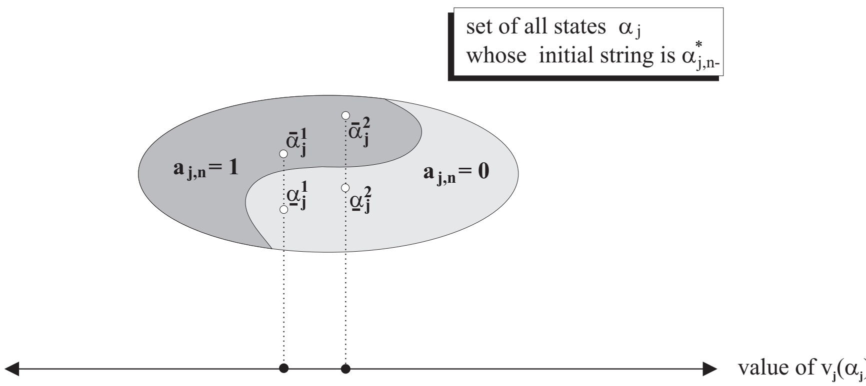

Figure 1: Reversibility of Contingency n for Agent j.

4 Reversibility and Strong Irreversibility

Reversibility of contingency n is a property of agent j's valuation function $v_j(\cdot)$ that can make incentive compatible revelation of $a_{j,n}$ bind as a constraint in the design of a finite contract f. Conversely, strong irreversibility of contingency n is a property of $v_j(\cdot)$ that can insure that incentive compatible revelation of $a_{j,n}$ does not constrain the design of f. Reversibility and strong irreversibility are thus useful in determining which contingencies can be addressed by incentive compatible finite contracts. These properties will be shown in subsequent sections to be significant in determining whether or not an optimal contract is recordable.

Definition 4 *Contingency n is **reversible** for agent j if for any initial string $\alpha_{j,n-}$ one can select at least two pairs of tails $(\underline{\alpha}^1_{j,n+}, \overline{\alpha}^1_{j,n+})$, $(\underline{\alpha}^2_{j,n+}, \overline{\alpha}^2_{j,n+})$ so that the following two properties hold.*

1. *Each pair of tails $(\underline{\alpha}^k_{j,n+}, \overline{\alpha}^k_{j,n+})$ perfectly reverses the effect upon agent j's valuation of contingency n: for $k = 1, 2$,*

$$v_j(\alpha_{j,n-}, 0, \underline{\alpha}^k_{j,n+}) = v_j(\alpha_{j,n-}, 1, \overline{\alpha}^k_{j,n+}). \tag{17}$$

2. *The pairs of tails differ in their effects upon agent j's valuation:*

$$v_j(\alpha_{j,n-}, 0, \underline{\alpha}^1_{j,n+}) \neq v_j(\alpha_{j,n-}, 0, \underline{\alpha}^2_{j,n+}). \tag{18}$$

Reversibility of contingency n is depicted in Figure 1, and the possibility that multiple types might determine the same valuation is illustrated in Example 2 of the last section. Definition 4 pinpoints a problem in incentive compatibility for finite contracts. To illustrate this point, consider an incentive compatible and finite contract f that is determined by the initial strings of length n. Suppose that the nth contingency observed by agent j is reversible. Agent j's interim expected utility given his type α_j is

$$H_j(\alpha_{j,n+1-})v_j(\alpha_j) + T_j(\alpha_{j,n+1-}),$$

where the tail $\alpha_{j,n+}$ beyond the nth contingency is omitted from $H_j(\cdot)$ and $T_j(\cdot)$ because f does not depend upon it. Because $v_j(\alpha_{j,n-}, 0, \underline{\alpha}^k_{j,n+}) = v_j(\alpha_{j,n-}, 1, \underline{\alpha}^k_{j,n+})$, it must be true that

$$\begin{aligned} H_j(\alpha_{j,n-}, 0)v_j(\alpha_{j,n-}, 0, \underline{\alpha}^k_{j,n+}) + T_j(\alpha_{j,n-}, 0) = \\ H_j(\alpha_{j,n-}, 1)v_j(\alpha_{j,n-}, 1, \underline{\alpha}^k_{j,n+}) + T_j(\alpha_{j,n-}, 1), \end{aligned} \tag{19}$$

or else agent j with type equal to either $(\alpha_{j,n-}, 0, \underline{\alpha}^k_{j,n+})$ or $(\alpha_{j,n-}, 1, \underline{\alpha}^k_{j,n+})$ would report whichever of these two types produced the larger of the two sides of this equation. This would contradict incentive compatibility for one of these two types. Now let $v \in \mathbb{R}$ denote a variable and consider the equation

$$[H_j(\alpha_{j,n-}, 0) - H_j(\alpha_{j,n-}, 1)] \cdot v = T_j(\alpha_{j,n-}, 1) - T_j(\alpha_{j,n-}, 0). \tag{20}$$

Statements (18) and (19) imply that (20) holds for distinct values of v, from which we conclude that

$$H_j(\alpha_{j,n-}, 0) = H_j(\alpha_{j,n-}, 1) \quad \textit{and} \quad T_j(\alpha_{j,n-}, 1) = T_j(\alpha_{j,n-}, 0)$$

for all $\alpha_{j,n-}$. Given our assumptions here on f, the two conditions (17) and (18) that define reversibility place conflicting incentive constraints upon an agent's interim expected utility function. The result is that the two terms $H_j(\cdot)$ and $T_j(\cdot)$ that capture the effect of the contract f upon interim expected utility cannot depend upon the reversible contingency. The reversible contingency can thus affect interim expected utility through agent j's valuation $v_j(\cdot)$ but not through its effect upon the contract f.

The above discussion shows that an incentive compatible finite contract f may be constrained in how it depends upon a reversible contingency. As will be shown, this constraint upon f can cause inefficiency. The following theorem presents a slightly different scenario in which the same conclusion holds; with an eye towards our later results, the assumption that f does not depend upon $\alpha_{j,n+}$ is replaced in this theorem with the assumption that each contingency in the tail is reversible.

Theorem 3 *For some $n \in \mathbb{N}$, suppose that every contingency observed by agent j after the nth is reversible. If the contract f is incentive compatible and finite, then the functions $H_j(\alpha_j)$ and $T_j(\alpha_j)$ determined by f depend only upon the first n contingencies. Consequently, agent j's interim expected utility in the contract f depends upon the tail $\alpha_{j,n+}$ only through its effect upon his valuation $v_j(\cdot)$.*

The proof is straightforward. Suppose that f does not depend on any contingency observed by agent j after the tth for some $t > n$. The argument above shows that $H_j(\alpha_j)$ and $T_j(\alpha_j)$ cannot depend upon $a_{j,t}$. The theorem then follows by backwards induction.

Theorem 3 begins our efforts to show how reversibility of a contingency constrains contract design, an idea that we will explore more deeply in the context of the principal-agent and bilateral trade models that follow. It is common in mechanism design to show that an incentive compatible mechanism can be constructed in a particular problem only by appropriately building inefficiency into the collective choice. As suggested by the above discussion, we will show that reversibility in these models can force the agents to design inefficiency into their contract by making it insensitive to certain contingencies as a means of achieving incentive compatibility.

Definition 5 *Contingency n is* ***strongly irreversible*** *for agent j if for every initial string $\alpha_{j,n-}$ there do not exist two pairs of tails $(\underline{\alpha}^1_{j,n+}, \overline{\alpha}^1_{j,n+})$, $(\underline{\alpha}^2_{j,n+}, \overline{\alpha}^2_{j,n+})$ that satisfy (17) and (18).*

"Strongly irreversible" is more demanding than "irreversible": contingency n is irreversible if there is *at least one* initial string $\alpha_{j,n-}$ for which no pair of tails exists satisfying (17) and (18), while contingency n is strongly irreversible if no such pair exists for *any* initial string $\alpha_{j,n-}$. We define "strongly irreversible" as above because it is useful in this form as a sufficient condition for proving that an optimal contract is recordable. Strong irreversibility is interpreted after the following example.

Example 3 *Let $v(\alpha) = \sum_{i=1}^{\infty} \delta^i a_i$ for $\delta \in (0,1)$. We show in this example that each of the properties of strong irreversibility and reversibility of contingency n holds for an interval of values of $\delta \in (0,1)$. This supports the hypothesis that neither of these two properties of $v(\cdot)$ is degenerate in our general model of contracting. Let α_{n-} be any initial string. The inequality*

$$\begin{aligned} & v(\alpha_{n-}, 0, \underline{\alpha}_{n+}) - v(\alpha_{n-}, 1, \overline{\alpha}_{n+}) \\ & \leq v(\alpha_{n-}, 0, 1, 1, \ldots) - v(\alpha_{n-}, 1, 0, 0, \ldots) \qquad (21) \\ & = \delta^n \left(\frac{2\delta - 1}{1 - \delta} \right) \qquad (22) \end{aligned}$$

holds for any two tails $\underline{\alpha}_{n+}, \overline{\alpha}_{n+}$. Reversibility or strong irreversibility of contingency n depends upon whether δ is at most or exceeds 0.5.

Case 1: $\delta \in (0, 0.5]$. *Each contingency is strongly irreversible. For arbitrary α_{n-} and $\delta = 0.5$, the right side of (22) equals 0 and (21) is strict except when $\underline{\alpha}_{n+} = (1, 1, \ldots)$ and $\overline{\alpha}_{n+} = (0, 0, \ldots)$. Condition (17) in the definition of reversibility thus holds only for this one choice of $\underline{\alpha}_{n+}$ and $\overline{\alpha}_{n+}$. For $\delta \in (0, 0.5)$ the right side of (21) is negative, which means that (17) can never hold.*

Case 2: $\delta \in (0.5, 1)$. *Each contingency is reversible. The right side of (21) is positive in this case. It is also true that*

$$v(\alpha_{n-}, 0, \underline{\alpha}_{n+}) - v(\alpha_{n-}, 1, \underline{\alpha}_{n+}) = -\delta^n.$$

Holding the initial string α_{n-} *constant,* $v(\alpha_{n-}, 0, \underline{\alpha}_{n+}) - v(\alpha_{n-}, 1, \overline{\alpha}_{n+})$ *covers an interval on the real line that contains* 0 *in its interior as* $\overline{\alpha}_{n+}$ *is varied. Choose* $\underline{\alpha}^1_{n+}$ *and* $\underline{\alpha}^2_{n+}$ *so that (18) holds. For each of these two tails* $\underline{\alpha}^k_{n+}$*, there exists a corresponding* $\overline{\alpha}^k_{n+}$ *so that (17) is satisfied, which verifies the reversibility of contingency* n*.*

Theorem 4 shows that strong irreversibility of each contingency implies an "orderliness" of an agent's valuation function $v(\cdot)$ with respect to initial strings. This property will be shown below to facilitate the approximation of a mechanism with a finite contract, which is the key step in our proofs of recordability of optimal contracts. Suppose that $A = \{0,1\}^{\mathbb{N}}$ and $v(A) = [\underline{v}, \overline{v}]$. If every contingency is strongly irreversible and $v(\cdot)$ is continuous, then Theorem 4 asserts the existence of a set of points

$$\underline{v} = x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(2^{n-1}+1)} = \overline{v} \tag{23}$$

such that the set of valuations determined by all types that share a particular initial string α_{n-} satisfies

$$v(\{\alpha_{n-}\} \times A_{n-1+}) = \left[x_{(k)}, x_{(k+1)}\right]$$

for some $1 \leq k \leq 2^{n-1}$. This is illustrated for $n = 2$ in case 1 of Example 2, where $0 = \underline{v} = x_{(1)}$, $x_{(2)} = 0.5$, and $x_{(3)} = 1$. As illustrated by cases 2 and 3 of Example 2, the interval of valuations is not partitioned according to the initial strings of a given length if contingencies are reversible: reversibility instead implies that the sets of valuations determined by distinct initial strings of the same length intersect nontrivially.

Theorem 4 *Assume that every contingency observed by an agent is strongly irreversible, that his valuation function* $v(\cdot)$ *is continuous on* $A = \{0,1\}^{\mathbb{N}}$*, and that* $v(A) = [\underline{v}, \overline{v}]$*. For any initial string* α_{n-}*, let* $D_{\alpha_{n-}} = v(\{\alpha_{n-}\} \times A_{n-1+})$*. Then the following statements hold.*

1. *Each set* $D_{\alpha_{n-}}$ *is a closed interval.*
2. *If* $\alpha_{n-} \neq \alpha'_{n-}$*, then* $D_{\alpha_{n-}} \cap D_{\alpha'_{n-}}$ *contains at most one point.*

The significance of Theorem 4 for our purposes is now explained in the case in which only one agent has private information. Applying this theorem, strong irreversibility of contingencies implies a one-to-one correspondence between finite contracts f that are determined by initial strings of length $n-1$ and mechanisms $\hat{f}$ that are constant on the intervals between the valuations in (23). The mechanism $\hat{f}$ is defined by step functions whose discontinuities occur at the valuations in (23), and the correspondence is through the composition $f = \hat{f} \circ v$. Continuity of $v(\cdot)$ implies that the intervals in (23) define an increasingly fine partition of $[\underline{v}, \overline{v}]$ as n increases. Any mechanism on $[\underline{v}, \overline{v}]$ that satisfies certain regularity properties can thus be approximated arbitrarily closely by a step function mechanism $\hat{f}$ of this kind. It is therefore approximated by the corresponding finite contract f. This is the insight that underlies our construction of sequences of finite contracts to prove the recordability of optimal contracts in Theorems 6 and 8 below.

5 Reversibility in a Principal-Agent Model

In the following two sections, we exploit the structure of two particular contracting problems to strengthen the conclusion of Theorem 3 concerning the effect of reversibility. Consider first a principal-agent problem with hidden information.[21] We begin by discussing this problem from the classical perspective that allows infiniteness of contracts and unlimited observability. The agent's privately observed type is $\alpha \in A = \{0,1\}^{\mathbb{N}}$. The agent makes a report to the principal concerning his type. The principal responds with an effort e to be made by the agent and a transfer t from the principal to the agent. The agent's effort e is observed by the principal. Let $e \in [0, \overline{e}]$ and $t \in [\underline{t}, \overline{t}]$. The principal's utility function is $u_0(e,t) = h(e) - t$, where $h(e)$ is a continuous and strictly increasing production function. The agent's utility function is $u_1(\alpha, e, t) = t + v(\alpha)e$, where $v(\alpha) < 0$ for all $\alpha \in A$.

A contract specifies an effort $e(\alpha)$ and a transfer $t(\alpha)$ as functions of the agent's type α. The principal-agent literature typically focuses on maximizing the ex ante expected utility of the principal, reflecting the idea that the principal selects the contract and offers it to the agent. A contract must provide the agent with a payoff of at least his reservation wage r and it must be in the agent's interest to honestly report α given that he will then take the action $e(\alpha)$. Incentive compatibility and interim individual rationality constraints must therefore be satisfied for the agent. The optimal contract for the principal solves

$$\max_{e(\cdot),t(\cdot)} \int h(e(\alpha)) - t(\alpha) d\pi(\alpha) \quad \textit{s.t. IC and IIR.} \tag{24}$$

This differs from the optimal mechanism problem in that the variable (e,t) in (24) can depend upon the state α and not just the valuation $v(\alpha)$. As discussed in section 2.3, however, a solution $(\hat{e}(v), \hat{t}(v))$ to the optimal mechanism problem determines a solution to (24) through composition with the agent's valuation function $v(\cdot)$. Reflecting the emphasis of the principal-agent literature on the ex ante utility of the principal, a contract (e,t) is *recordable* if there exists a sequence of finite contracts $(e_n, t_n)_{n \in \mathbb{N}}$ satisfying *IC* and *IIR* such that

$$\lim_{n\to\infty} \int h(e_n(\alpha)) - t_n(\alpha) d\pi(\alpha) \geq \int h(e(\alpha)) - t(\alpha) d\pi(\alpha).$$

[21] See, for instance, Mas-Colell, Whinston, and Green (1995, pp. 900–903).

The following theorem addresses the scenario considered in Theorem 3 in which each contingency n observed by the agent is reversible for $n \geq r$. Statement 1 applies Theorem 3 to conclude that an incentive compatible finite contract depends only upon the first $r-1$ contingencies.[22] This result is significant because it identifies an infinite set of contingencies that cannot be addressed by an incentive compatible finite contract. Statements 2 and 3 are discussed below.

Theorem 5 *If there exists $r \in \mathbb{N}$ such that every contingency $n \geq r$ is reversible for the agent, then the following statements hold.*

1. *An incentive compatible finite contract (e^*, t^*) can only depend on a_n for $n < r$.*

2. *A recordable contract (e, t) is weakly ex ante dominated for the principal by an incentive compatible, interim individually rational finite contract (e^*, t^*), i.e.,*

$$\int h(e^*(\alpha)) - t^*(\alpha) d\pi(\alpha) \geq \int h(e(\alpha)) - t(\alpha) d\pi(\alpha).$$

3. *If only infinite contracts solve the optimal contracting problem (24), then optimal contracts are not recordable.*

Statement 2 can be regarded as a lemma that sets up statement 3. Statement 2 asserts that the ex ante expected return of the principal in any recordable contract is weakly dominated by his expected return in some incentive compatible, interim individually rational finite contract. If an optimal contract is recordable, then statement 2 implies that some finite contract is optimal. The conclusion in statement 3 concerning a problem in which only infinite contracts can be optimal thus follows immediately from statement 2.

Statement 3 is significant because only infinite contracts can be optimal given standard regularity conditions on the distribution μ of the agent's valuation.[23] Reversibility of all contingencies in a tail therefore causes contractual incompleteness in the most commonly studied version of this principal-agent model.

[22] The functions $t(\alpha)$ and $e(\alpha)$ correspond to the expected values $T_1(\alpha_1)$ and $H_1(\alpha_1)$, respectively, in Theorem 3. The one-sided uncertainty of the principal-agent model thus allows the independence of the tail of reversible contingencies to be strengthened here from an interim to an ex post result. This is not true in the bilateral trade model of section 6 because of its two-sided incomplete information.

[23] These regularity conditions can be found in Mirrlees (1971) or Fudenberg and Tirole (1991). The argument is as follows. If $(e(\alpha), t(\alpha))$ is an finite contract that is optimal, then the mechanism $(\hat{e}(v), \hat{t}(v))$ defined from $(e(\alpha), t(\alpha))$ in the proof of Theorem 11 is an optimal mechanism. The finiteness of $(e(\alpha), t(\alpha))$ together with the definition of $(\hat{e}(v), \hat{t}(v))$ imply that the functions $\hat{e}(v)$ and $\hat{t}(v)$ assume only a finite number of values. The conditions on μ, however, imply that the functions $\hat{e}(v)$ and $\hat{t}(v)$ in an optimal mechanism are strictly increasing over $[\underline{v}, \overline{v}]$ and therefore assume an infinite number of values. This contradiction implies that only infinite contracts solve (24) when these regularity conditions hold.

Conversely, Theorem 6 shows that the optimal contract can be recordable if every contingency is strongly irreversible. The regularity condition 3 in this theorem is needed so that the standard characterization of the optimal mechanism can be applied. The theorem is proven by constructing a sequence of finite contracts to demonstrate recordability of the optimal contract. The construction follows the method discussed immediately after the statement of Theorem 4.

Theorem 6 *Assume that:*

1. *each contingency is strongly irreversible;*
2. $v(\cdot)$ *is continuous;*
3. $v(A) = [\underline{v}, \overline{v}]$, $\overline{v} < 0$, *the induced distribution* μ *is nonatomic, and* μ *has* $[\underline{v}, \overline{v}]$ *as its support;*
4. *for all* $n \in \mathbb{N}$, *every initial string* $\alpha_{n-} \in A_{n-}$ *occurs with positive probability;*
5. *an optimal mechanism* $(\hat{e}, \hat{t})$ *exists such that* $\hat{e}$ *and* $\hat{t}$ *are continuous on* $[\underline{v}, \overline{v}]$.[24]

Let (e, t) *be an optimal contract such that* $e = \hat{e}(v(\cdot))$ *and* $t = \hat{e}(v(\cdot))$. *Then* (e, t) *is recordable. The sequence of contracts that demonstrates recordability can be chosen to converge uniformly, and hence the efforts and transfers of* (e, t) *are approximated by those in the sequence.*

This section concludes with four examples. Example 4 concerns the construction of a sequence of finite contracts in the proof of Theorem 6 to demonstrate the recordability of an optimal contract. This example concerns the case of

$$v(\alpha) = \overline{v} - \frac{1-\delta}{\delta} \sum_{t=1}^{\infty} \delta^t a_t \tag{25}$$

where $\overline{v} < 0$ is some constant and $\delta \in [0.5, 1)$.[25] Case 1 of this example describes the successful construction of the sequence when each contingency is strongly irreversible ($\delta = 0.5$), while case 2 illustrates why the construction fails when each contingency is reversible ($\delta \in (0.5, 1)$). Examples 5-7 explore strong irreversibility as a principle for identifying the contingencies that may be addressed by incentive compatible finite contracts. While our work on this topic is clearly not comprehensive, it is a initial step towards understanding the properties of agents' preferences over states of the world that determine the potential content of contracts.

[24] A sufficient condition for continuity of the optimal contract is that the hazard rate μ'/μ is increasing (e.g., Fudenberg and Tirole (1991, p. 267)).

[25] The number $\overline{v}$ insures that $v(\alpha) < 0$, i.e., there is always cost associated with effort. This insures that a finite effort is optimal. The role of number $\overline{v}$ in Examples 4, 6 and 7 is solely to translate all intervals leftward so that $v(\cdot)$ is negative.

Example 4 *Strong irreversibility is used in the proof of Theorem 6 to show that a finite contract corresponds to a step function mechanism and, conversely, a step function mechanism whose steps are at appropriately chosen points corresponds to a finite contract. We illustrate this point with formula (25) for* $v(\cdot)$ *and* $\delta \in [0.5, 1)$. *Notice that* $v(A) = \overline{v} + [-1, 0]$.

Case 1: $\delta = 0.5$. *Case 1 of Example 3 shows that each contingency* n *is strongly irreversible and hence Theorem 6 applies. For any initial string* α^*_{n+1-} *of length* n,

$$v\left(\{\alpha^*_{n+1-}\} \times A_{n+1+}\right) = \overline{v} + \left[-k\left(\frac{1}{2}\right)^n, -(k-1)\left(\frac{1}{2}\right)^n\right] \tag{26}$$

for some integer $1 \leq k \leq 2^n$. *Varying* α^*_{n+1-} *over* A_{n+1-} *produces the* 2^n *intervals of the form (26). To prove Theorem 6 in this case, construct an incentive compatible and individually rational mechanism* $(e^*_n(v), t^*_n(v))$ *that "approximates" the optimal mechanism and with the property that* $e^*_n(v)$ *and* $t^*_n(v)$ *are constant on the interior of each interval of the form (26). By taking some care with the definition of* $(e^*_n(v(\alpha)), t^*_n(v(\alpha)))$ *at values of* α *that are mapped by* $v(\cdot)$ *into the endpoints of the intervals in (26), this contract will depend only upon the initial string of length* n. *Finer approximations can be constructed as* n *increases, which is how a sequence that converges to the optimal contract is constructed in the proof.*

Case 2: $\delta \in (0.5, 1)$. *Case 2 of Example 3 shows that each contingency* n *is reversible and hence Theorem 6 does not apply. What goes wrong in the construction? Suppose that* $(e^*_n(v), t^*_n(v))$ *is a mechanism with the property that* $(e^*_n(v(\alpha)), t^*_n(v(\alpha)))$ *depends only upon the initial string* α_{n+1-} *of length* n *for all* $\alpha_{n+1-} \in A_{n+1-}$. *The finiteness of* $(e^*_n(v(\alpha)), t^*_n(v(\alpha)))$ *implies that* $e^*_n(v)$ *and* $t^*_n(v)$ *are constant on each interval of the form*

$$v\left(\{\alpha_{n+1-}\} \times A_{n+1+}\right) = \overline{v} + \left[-\sum_{t=1}^{n} a_t\delta^t - \delta^t, -\sum_{t=1}^{n} a_t\delta^t\right].$$

The analysis of case 3 of Example 2 implies that (unlike case 1 above) the interiors of these intervals cover $\overline{v}+(-1, 0)$. *The connectedness of* $\overline{v}+(-1, 0)$ *then implies that* $e^*_n(v)$ and $t^*_n(v)$ *must be constant on this interval. Approximating the optimal mechanism with* $(e^*_n(v), t^*_n(v))$ *and then considering the induced contract* $(e^*_n(v(\alpha)), t^*_n(v(\alpha)))$ *is therefore ineffective as a technique here because only constant mechanisms can be used in the approximation.*

This last argument does not address the possibility of constructing contracts by other methods that approximate the optimal contract. Theorem 3 (as applied in statement 1 of Theorem 5) demonstrates that other approaches will also be unsuccessful. This general result depends crucially upon the constraint of incentive compatibility, while the argument in the specific case above does not.

Example 5 *A firm's central administration (the principal) contracts with a division (the agent) for a service. A contingency is any binomial random* *variable that can be observed by the division whose realization affects its cost of providing the service to the center. Suppose that the first contingency represents the ability (high or low) of a particular employee. This employee is a key person to the division if the effect of his ability upon the division's cost of providing the service cannot be undone or masked by the realizations of other contingencies.*[26] *Formally, let* $v(\alpha)$ *be the division's cost function given realizations* $\alpha = (a_n)_{n\in\mathbb{N}}$ *of the contingencies. Employee 1 is key if there do not exist tails* $\underline{\alpha}_{1+}, \overline{\alpha}_{1+}$ *such that* $v(0, \underline{\alpha}_{1+}) = v(1, \overline{\alpha}_{1+})$. *Employee 1 being a key person is thus sufficient to insure that contingency* a_1 *representing his ability is strongly irreversible. A thorough analysis of a cost function of this kind can be found in Examples 6 and 7 below. A key person is distinguished not solely by the magnitude of the effect of his ability upon cost but also its relation to the aggregate effect of all other contingencies. Our theory suggests that the contract between the principal firm and the agent firm can address the key person's ability. This is consistent with common practice, where contracts between firms commonly address the participation of key people.*[27]

Example 6 *We examine the impact of the first contingency upon the valuation function by setting*

$$v(\alpha) = \overline{v} - \left(a_1 V + \left(\frac{1-\delta}{\delta^2} \right) \sum_{t=2} a_t \delta^t \right),$$

where the values of $V > 0$ *and* $\delta \in [0.5, 1)$ *are specified below. At issue here is (i) whether or not the optimal contract is recordable, and (ii) if it is not, then whether or not contingency 1 can be addressed in an incentive compatible finite contract. The term* $(1-\delta)/\delta^2$ *normalizes the sum so that*

$$\begin{aligned} v(\{a_1 = 0\} \times A_{1+}) &= \overline{v} + [-1, 0] \quad \text{and} \\ v(\{a_1 = 1\} \times A_{1+}) &= \overline{v} + [-V-1, -V]. \end{aligned} \tag{27}$$

Recall that Example 3 shows that contingency n *for* $n \geq 2$ *is strongly irreversible if* $\delta = 0.5$ *and reversible if* $\delta \in (0.5, 1)$. *We begin here by showing that contingency 1 is strongly irreversible if* $V \geq 1$ *and reversible if* $V \in (0, 1)$. *If* $V \geq 1$, *then (27) implies that* $v(0, \underline{\alpha}_{1+}) = v(1, \overline{\alpha}_{1+})$ *can hold only if* $V = 1$, $\underline{a}_t = 1$, *and* $\overline{a}_t = 0$ *for all* $t \geq 2$. *Equation (17) in the definition of reversibility therefore can hold for at most one pair of tails* $(\underline{\alpha}_{1+}, \overline{\alpha}_{1+})$. *Because the set of strings of length zero is vacuous, this establishes strong irreversibility*

[26] The term "key person" is drawn from the insurance industry, where "key person insurance" is a form of term life insurance sold to a firm to protect itself against the loss through death or disability of an individual who is crucial to the success of the firm. The issuance of such a policy requires not only that the health of the individual be appraised but also that his value to the firm be documented. The insured value of such an individual may run in the tens of millions of dollars.

[27] Loan contracts, for instance, commonly require that the borrowing firm adequately insure against loss through disability or death of key employees. This is a completely separate issue from the incentives that the borrowing firm may provide its own employees to motivate their performance, which is the usual focus of the principal-agent literature.

of contingency 1 for $V \geq 1$. If $V < 1$, the intersection of the two intervals in (27) has a nonempty interior. Choose $v' \neq v''$ in this interior. There exists tails $\underline{\alpha}'_{1+}, \overline{\alpha}'_{1+}, \underline{\alpha}''_{1+}, \overline{\alpha}''_{1+} \in A_{1+}$ such that

$$v\left(0, \underline{\alpha}'_{1+}\right) = v\left(1, \overline{\alpha}'_{1+}\right) = v' \neq v'' = v\left(0, \underline{\alpha}''_{1+}\right) = v\left(1, \overline{\alpha}''_{1+}\right),$$

which shows that contingency 1 is reversible.

We now address issues (i) and (ii) in the four cases determined by the values of V and δ.

Case 1: *$V \geq 1$ and $\delta = 0.5$. All contingencies are strongly irreversible. If its regularity conditions hold, then Theorem 6 implies that the optimal contract is recordable.*

Case 2: *$V < 1$ and $\delta \in (0.5, 1)$. All contingencies are reversible. Theorem 5 implies that any incentive compatible contract is independent of the state α. In particular, it cannot address contingency 1.*

Case 3: *$V \geq 1$ and $\delta \in (0.5, 1)$. Theorem 5 implies that an incentive compatible finite contract may depend at most upon a_1. The optimal contract is thus not recordable.*

Case 4: *$V < 1$ and $\delta = 0.5$. Contingency 1 is reversible but all subsequent contingencies are not. Our theorems do not explicitly address this case.*

Example 7 *Cases 3 and 4 of the last example leave open the possibility that an incentive compatible contract might exist that only addresses contingency 1 in these cases. Reflecting the role of reversibility in determining whether or not a contingency can be contracted upon, we now show that such a contract exists in case 3 in which contingency 1 is strongly irreversible but not in case 4 in which it is reversible. Consider*

$$e(a_1) = \begin{cases} e' & \text{if } a_1 = 0 \\ e'' & \text{if } a_1 = 1 \end{cases} \quad \text{and} \quad t(a_1) = \begin{cases} t' & \text{if } a_1 = 0 \\ t'' & \text{if } a_1 = 1 \end{cases}.$$

The issue is whether or not e', e'', t', t'' can be chosen so that (i) (e, t) is incentive compatible and (ii) $e' \neq e''$ or $t' \neq t''$ so the contract depends upon a_1. The contract (e, t) is incentive compatible if and only if

$$t' + v^* e' \geq t'' + v^* e'' \tag{28}$$

for all $v^* \in v(\{a_1 = 0\} \times A_{1+}) = \overline{v} + [-1, 0]$, and

$$t'' + v^{**} e'' \geq t' + v^{**} e' \tag{29}$$

for all $v^{**} \in v(\{a_1 = 1\} \times A_{1+}) = \overline{v} + [-V - 1, -V]$. Assuming $e' \neq e''$ or $t' \neq t''$, these inequalities cannot hold for all

$$v^* = v^{**} \in \overline{v} + Int([-1, 0] \cap [-V - 1, -V]),$$

and hence this intersection must have an empty interior. An incentive compatible contract thus cannot depend upon contingency a_1 alone if $V \in (0,1)$ (i.e, a_1 is reversible). If $V \geq 1$, then (28)-(29) hold if and only if e', e'', t', t'' satisfy

$$v^* \left(e'' - e'\right) \leq t' - t'' \leq v^{**} \left(e'' - e'\right) \tag{30}$$

for all $v^* \in \overline{v} + [-1, 0]$ and $v^{**} \in \overline{v} + [-V - 1, -V]$. If $V \geq 1$, then (30) holds if $e' - e'' > 0$ and

$$\left(\overline{v} - 1\right)\left(e'' - e'\right) \leq t' - t'' \leq \left(\overline{v} - V\right)\left(e'' - e'\right).$$

A contract that addresses just a_1 can therefore be incentive compatible if and only if $V \geq 1$ (i.e., a_1 is strongly irreversible).

6 Reversibility in a Model of Bilateral Trade

We next consider a contracting problem with two-sided incomplete information. A seller can provide a service to a buyer. The state of the world specifies every detail that affects the value of the service to the buyer and the cost of provision to the seller.[28] The buyer's and the seller's types are $\alpha_B = (a_{B,i})_{i \in \mathbb{N}} \in A_B$ and $\alpha_S = (a_{S,i})_{i \in \mathbb{N}} \in A_S$, respectively, with π_B and π_S denoting the distributions of these types. A *contract* is a pair (p, t) that specifies for each α_B and α_S a probability $p(\alpha_B, \alpha_S)$ that the seller provides the service to the buyer and a transfer $t(\alpha_B, \alpha_S)$ from the buyer to the seller. Contracts are thus assumed to be ex post budget balanced throughout this discussion. The buyer's utility is

$$u_B(\alpha_B, p, t) = p \cdot v_B(\alpha_B) + t$$

and the seller's utility is

$$u_S(\alpha_S, p, t) = t - p \cdot v_S(\alpha_S),$$

where $v_B : A_B \rightarrow [\underline{v}_B, \overline{v}_B] \subset \mathbb{R}^+$ and $v_S : A_S \rightarrow [\underline{v}_S, \overline{v}_S] \subset \mathbb{R}^+$. A contract is required to satisfy *IC* for both the buyer and the seller because each trader has private information. A contract is required to satisfy *IIR* for each trader given that his default utility is 0. It is assumed throughout this section that any initial string observable by either trader occurs with positive probability: for $j = B, S$,

$$\pi_j\left(\alpha_{j,n-}\right) = \pi_j\left(\{\alpha_{j,n-}\} \times A_{j,n-1+}\right) > 0 \tag{31}$$

[28] For example, consider a service provider and a system owner who negotiate a service contract for a computer system. The contingencies of the owner capture all aspects of his business that determine the value he would receive from a properly maintained system, including future decisions by current and prospective customers and employees. For the service provider, the contingencies include everything that affects his cost of providing the service, including service demands from other customers, the difficulty of repairing different components of the system, and whether or not the various components require service at the interim stage.

for any $n \in \mathbb{N}$ and $\alpha_{j,n-} \in A_{j,n-}$.

Let μ_B denote the distribution of the buyer's valuation defined by v_B and π_B, and let μ_S denote the distribution of the seller's valuation defined by v_S and π_S. The model of bilateral trade of Chatterjee and Samuelson (1983) is a special case of our model in which the densities μ'_B and μ'_S are continuous and nonzero on $[\underline{v}_B, \overline{v}_B]$ and $[\underline{v}_S, \overline{v}_S]$, respectively. Our approach extends their model by modeling as states of the world those aspects of the service or good that determine the payoffs from trading.[29]

The *optimal contract* (p^*, t^*) maximizes the expected gains from trade subject to *IC* and *IIR*: (p^*, t^*) solves

$$\max_{(p,t)} \int_A [v_B(\alpha_B) - v_S(\alpha_S)]\, p(\alpha) d\pi(\alpha) \quad \textit{s.t. IC and IIR.} \tag{32}$$

As in the principal-agent problem, Theorem 11 reduces the problem of designing optimal contracts to the problem of designing optimal mechanisms. Subject to some regularity conditions on μ_B and μ_S, the optimal mechanism is determined in Myerson and Satterthwaite (1983).

Reflecting the objective in (32) of maximizing the expected gains from trade, a contract (p, t) is *recordable* if there exists a sequence of finite contracts $(p_m, t_m)_{m \in \mathbb{N}}$ satisfying *IC* and *IIR* such that

$$\lim_{m \to \infty} \int_A [v_B(\alpha_B) - v_S(\alpha_S)]\, p_m(\alpha) d\pi(\alpha) \geq \int_A [v_B(\alpha_B) - v_S(\alpha_S)]\, p(\alpha) d\pi(\alpha). \tag{33}$$

Our first two theorems concerning bilateral trade mirror the two results of section 5 concerning the principal-agent model. Suppose that all but a finite number of contingencies observed by each trader are reversible. Given a recordable contract, Theorem 7 states that there exists an incentive compatible, interim individually rational finite contract that achieves at least as much of the potential gains from trade as the given recordable contract. As in Theorem 5, Theorem 7 thus implies that optimal contracts are not recordable in problems in which only infinite contracts can be optimal. The Myerson and Satterthwaite (1983) characterization of the optimal mechanism suggests that optimal contracts are typically infinite. Theorem 7 in this sense provides sufficient conditions for contractual incompleteness.

Theorem 7 *For some $b, s \in \mathbb{N}$, suppose that a contingency $a_{B,n}$ with $n \geq b$ and a contingency $a_{S,n}$ with $n \geq s$ is reversible for the trader who observes its realization. If the contract (p, t) is recordable, then there exists an incentive compatible, interim individually rational finite contract (p^*, t^*) such that:*

1. *$(p^*(\alpha), t^*(\alpha))$ does not depend upon $a_{B,n}$ for $n \geq b$ and $a_{S,n}$ for $n \geq s$;*

[29] We thus addresses a common weakness of noncooperative bargaining theory, which is that bargainers negotiate only price and perhaps quantity in common models, whereas real bargaining problems commonly concern a multitude of issues.

2. *the ex ante gains from trade in* (p^*, t^*) *are at least as large as in* (p, t), *i.e.,*

$$\int_A [v_B(\alpha_B) - v_S(\alpha_S)]\, p^*(\alpha) d\pi(\alpha) \geq \int_A [v_B(\alpha_B) - v_S(\alpha_S)]\, p(\alpha) d\pi(\alpha);$$

3. *Consequently, if only infinite contracts can be optimal, then an optimal contract is not recordable.*

Similar to Theorem 6, the next theorem concerns a case in which each contingency is strongly irreversible and the optimal contract is both infinite and recordable. This theorem applies the Myerson-Satterthwaite (1983) characterization of the optimal mechanism, which requires regularity conditions on μ_B and μ_S that are stated in assumptions 4 and 5 of the theorem. Suppose that the densities μ_B' and μ_S' exist, are continuous, and have $[0, 1]$ as their common support. For $k \in [0, 1]$, let $V_B(\cdot)$ and $V_S(\cdot)$ denote the *virtual valuation functions*

$$V_B(v_B, k) = v_B + k \frac{\mu_B(v_B) - 1}{\mu_B'(v_B)}, \quad (34)$$

$$V_S(v_S, k) = v_S + k \frac{\mu_S(v_S)}{\mu_S'(v_S)}. \quad (35)$$

Theorem 8 *Assume that:*

1. *each contingency is strongly irreversible;*

2. $v_B(\cdot)$ *and* $v_S(\cdot)$ *are continuous;*

3. *the distributions* μ_B *and* μ_S *have continuous, strictly positive densities on* $[0, 1]$*;*

4. *the virtual valuation functions* $V_B(\cdot, k)$ *and* $V_B(\cdot, k)$ *are increasing on* $[0, 1]$ *for each choice of* k.

Then an optimal contract (p^*, t^*) *is recordable. The sequence of contracts* $((p_n, t_n))_{n \in \mathbb{N}}$ *that demonstrates the recordability of* (p^*, t^*) *can be chosen so that* $\lim_{n \to \infty} p_n(\alpha) = p^*(\alpha)$ *except on a set of states of* π*-measure zero.*

Our final theorem applies Theorem 7 to reveal the severe contractual incompleteness when (i) every contingency is reversible, and (ii) it is not common knowledge ex ante that trade should occur. It is shown in this case that no gains from trade are achieved ex ante in any recordable contract. Ex ante contracting is thus pointless in this setting.

Theorem 9 *Assume that:*

1. *each contingency observed by a trader is reversible;*

2. *there exists types* α_B^* *and* α_S^* *for the buyer and the seller such that* $v_B(\alpha_B^*) < v_S(\alpha_S^*)$*, so that trade should not occur for these types.*

Then the ex ante expected gains from trade are zero in any recordable contract.

Example 8 *Theorem 9 is a rather striking result and so it is worthwhile to examine a well-known fixed-price game in relation to this theorem.*[30] *Assume that* $[\underline{v}_B, \overline{v}_B] = [\underline{v}_S, \overline{v}_S] = [0,1]$ *and fix a price* $\rho \in (0,1)$*. The game operates as follows: with access to his type, each trader announces "yes" or "no", and trade occurs at the price of* ρ *if and only if each trader announces "yes". This is inspired by the fixed price game of Hagerty and Rogerson (1985), with the minor change here that each trader announces "yes" or "no" rather than his valuation so that communication is limited. If each trader knows his valuation at the interim, then it is a dominant strategy for each trader to announce "yes" if and only if he can profitably trade at the price of* ρ*. Trade occurs in this equilibrium for* (v_B, v_S) *satisfying* $v_S \leq \rho \leq v_B$ *and so there are gains from trade.*

Let

$$v_B = \frac{1-\delta}{\delta}\sum_{t=1}^{\infty}\delta^t a_{B,t} \text{ and } v_S = \frac{1-\delta}{\delta}\sum_{t=1}^{\infty}\delta^t a_{S,t}$$

for $\delta \in [0.5, 1)$*. Case 1 below assumes that all contingencies are strongly irreversible for each trader* $(\delta = 0.5)$*. It is shown that the dominant strategy equilibrium for the fixed price of* $\rho = 0.5$ *defines a Bayesian-Nash equilibrium with limited observability. Case 2 assumes that all contingencies are reversible for each trader* $(\delta \in (0.5, 1))$*. It is shown in this case that the dominant strategy equilibrium determined by* ρ *does not define a Bayesian-Nash equilibrium with limited observability, regardless of the choice of* ρ*. As an illustration of the no-trade result of Theorem 9, it is then shown that any Bayesian-Nash equilibrium with limited observability in the case of reversible contingencies is the same as a particular no-trade equilibrium in almost all states* α*.*

Case 1: $\delta = 0.5$*. Fix the price at* $\rho = 0.5$*. The buyer knows whether or not* v_B *is as big or as small as* ρ *based solely upon observing his first contingency* $a_{B,1}$*. Formally,*

$$\begin{aligned} a_{B,1} &= 0 \Rightarrow v_B \leq 0.5, \text{ and} \\ a_{B,1} &= 1 \Rightarrow v_B \geq 0.5. \end{aligned}$$

The strategy

$$\gamma_B(\alpha_B) = \begin{cases} \text{no if } a_{B,1} = 0 \\ \text{yes if } a_{B,1} = 1 \end{cases}$$

[30] This example was suggested by a question from Jim Peck.

is finite and maximizes the buyer's expected payoff against any strategy of the seller, regardless of how long of an initial string the buyer may observe. Similar remarks apply to the seller. The trading outcome

$$p\left(v_B, v_S\right) = \begin{cases} \textit{1 if } v_S \leq 0.5 \leq v_B \\ \textit{0 otherwise} \end{cases}$$

with a transfer of 0.5 *if and only if trade occurs is therefore sustainable as a Bayesian-Nash equilibrium with limited observability. The corresponding contract*

$$\begin{aligned} p^*\left(\alpha_B, \alpha_S\right) &= \begin{cases} \textit{1 if } a_{S,1} = \textit{0 and } a_{B,1} = 1 \\ \textit{0 otherwise} \end{cases}, \textit{ and} \qquad (36) \\ t^*\left(\alpha_B, \alpha_S\right) &= \begin{cases} \textit{0.5 if } a_{S,1} = \textit{0 and } a_{B,1} = 1 \\ \textit{0 otherwise} \end{cases} \end{aligned}$$

is incentive compatible, interim individually rational and finite.

Case 2: $\delta \in (0.5, 1)$. *Let the price* ρ *be any fixed value in* $(0,1)$. *Case 3 of Example 2 implies that each contingency observed by a trader is reversible, and hence Theorem 9 together with the revelation principle of Theorem 2 imply that the expected gains from trade are zero in any Bayesian-Nash equilibrium with limited observability. A trader can of course determine conditional upon a finite string of length* $n-1$ *whether his expected valuation is no more than or at least* ρ. *At first glance, the finite strategies*

$$\begin{aligned} \gamma_B\left(\alpha_B\right) &= \begin{cases} \textit{no if } E\left[v_B \left|\alpha_{B,n-}\right.\right] < \rho \\ \textit{yes if } E\left[v_B \left|\alpha_{B,n-}\right.\right] \geq \rho \end{cases}, \textit{ and} \qquad (37) \\ \gamma_S\left(\alpha_S\right) &= \begin{cases} \textit{yes if } E\left[v_S \left|\alpha_{S,n-}\right.\right] \leq \rho \\ \textit{no if } E\left[v_S \left|\alpha_{S,n-}\right.\right] > \rho \end{cases} \end{aligned}$$

may seem to sustain positive expected gains from trade in equilibrium. These strategies, however, do not define a Bayesian-Nash equilibrium with limited observability, for a trader may wish to deviate from his specified message after observing a finite number of additional contingencies beyond the $(n-1)$*st. This can be formalized as follows. It is shown in the analysis of case 3 of Example 2 that there exists* $\alpha^*_{B,n-}$ *such that* $v_B(\{\alpha^*_{B,n-}\} \times A_{B,n+})$ *is a closed interval with* ρ *in its interior. Consequently, there exists* $t > n$ *and* $\alpha'_B, \alpha''_B \in A_B$ *such that*

$$\alpha'_{B,n-} = \alpha''_{B,n-} = \alpha^*_{B,n-},$$

but

$$\begin{aligned} E\left[v_B\left(\alpha_B\right) \left|\alpha_{B,t-} = \alpha'_{B,t-}\right.\right] &< \rho, \textit{ and} \\ E\left[v_B\left(\alpha_B\right) \left|\alpha_{B,t-} = \alpha''_{B,t-}\right.\right] &> \rho. \end{aligned}$$

Given the seller's use of $\gamma_S(\cdot)$, *the buyer's unique best response is to report "no" if he observes* $\alpha'_{B,t-}$ *and "yes" if he observes* $\alpha''_{B,t-}$. *The single report* $\gamma_B(\alpha'_B) = \gamma_B(\alpha''_B) = \gamma_B(\alpha^*_B)$ *is not a best response in each of these two instances, which means that* (γ_B, γ_S) *is not a Bayesian-Nash equilibrium with limited observability.*

The strategies

$$\begin{aligned} \gamma^*_B(\alpha_B) &= \text{no for all } \alpha_B \in A_B \text{ and} \\ \gamma^*_S(\alpha_S) &= \text{no for all } \alpha_S \in A_S \end{aligned}$$

define a Bayesian-Nash equilibrium with limited observability in which trade never occurs. An equilibrium thus exists for every $\delta \in (0.5, 1)$ *and* $\rho \in (0, 1)$. *We now show that* (γ^*_B, γ^*_S) *is the only equilibrium, except for trivial variations of* γ^*_B *and* γ^*_S *over sets of measure zero. The argument is by contradiction. Suppose that* (γ'_B, γ'_S) *is a Bayesian-Nash equilibrium with limited observability that differs from* (γ^*_B, γ^*_S) *on some set of states of positive* π*-measure. This means that at least one of the traders reports "yes" with positive probability in the equilibrium* (γ'_B, γ'_S). *Suppose for notational convenience that this is true of the seller; the contradiction is derived by considering the buyer's best response to* γ'_S. *The finiteness of the strategies in a Bayesian-Nash equilibrium with limited observability implies that there exists* $n \in \mathbb{N}$ *such that* $\gamma'_B(\alpha_B)$ *is determined by* $\alpha_{B,n-}$ *for all* $\alpha_{B,n-} \in A_{B,n-}$. *Because the seller's use of* γ'_S *presents the buyer with the positive probability of profitable trade,* $\gamma'_B(\alpha_B)$ *must equal* $\gamma_B(\alpha_B)$ *as defined in (37). The argument made above now provides the desired contradiction: for the initial strings* $\alpha'_{B,t-}, \alpha''_{B,t-}$ *and* $\alpha^*_{B,n-}$ *defined above, the buyer would deviate from* $\gamma'_B(\alpha^*_{Bn-})$ *upon observing one of* $\alpha'_{B,t-}$ *or* $\alpha''_{B,t-}$. *The pair* (γ'_B, γ'_S) *is thus not a Bayesian-Nash equilibrium with limited observability, which completes the contradiction.*

We conclude this discussion of bilateral trade by connecting our analysis to the broader aims of models of incomplete contracts. Results that show that agents may accomplish more in a later stage than they can achieve ex ante in a contract are desirable in a model of contractual incompleteness, first because they demonstrate the nonnegligible losses that make contractual incompleteness compelling as an issue, and second because they demonstrate the activity in the interim and ex post stages through which the agents try to accomplish what they failed to achieve ex ante. Renegotiation exemplifies this later activity. That such activity occurs in reality and is inconsistent with complete contracts is the principle issue that motivates the theory of contractual incompleteness. We now discuss the possibility that the buyer and the seller may trade at an interim state despite the absence of an ex ante contract that organizes their trading. Consider the fixed price game of Example 8 with the price of ρ and assume that specific types α^*_B and α^*_S of the traders are realized. Unless $v_B(\alpha^*_B)$ or $v_S(\alpha^*_S)$ equals ρ, continuity of each trader's valuation function insures that each trader can determine after observing a sufficiently large but finite number of contingencies whether his valuation is more or less than ρ. If each trader uses his dominant strategy of price-taking, then trade occurs if $v_B(\alpha_B^*) > \rho > v_S(\alpha_S^*)$.

It is thus arguable that profitable trade may occur in an interim state (α_B^*, α_S^*), even in the case of reversible contingencies in which gains from trade can not be contracted ex ante.[31] This stands in stark contrast to the "no information-based trading in equilibrium" result of Milgrom and Stokey (1982) or, more generally, the result that ex ante efficiency implies interim efficiency of Hölmstrom and Myerson (1983, p. 1806). Both of these classic results depend crucially upon the assumption that complete contracts are available to the agents ex ante. The primary difference between our results and these results is our constraint of limited observability and its implication that contracts are necessarily finite. As captured by Theorem 1, the fact that limited observability is more severe as a constraint ex ante than at the interim explains why ex ante performance may be worse in our model than interim performance.

7 Conclusion

Contractual incompleteness is inefficiency of a nonnegligible magnitude that occurs because a contract fails to address some contingencies. We develop a model in which all contingencies are foreseen by the agents in the sense that they know the structure of the model and yet contractual incompleteness may necessarily occur. The causes of contractual incompleteness in our model are: (i) limited observability, which is the inability of agents to observe more than a finite number of contingencies; (ii) a set of states of the world that is complex in the sense that there are an infinite number of possible contingencies; (iii) incomplete information and the consequent problem of incentive compatibility; (iv) reversibility of contingencies, which is a property of an agent's preferences over states of the world. Conversely, we identify strong irreversibility as a property of preferences under which contractual incompleteness need not occur in the context of (i)-(iii) in the sense that the optimal contract is recordable.

We emphasize in this paper the properties of preferences over states of the world that determine whether or not contractual incompleteness must occur. We explore in several results and also in Examples 5-8 attributes of contingencies that determine whether or not those contingencies can be successfully addressed in a contract. The task of identifying aspects of the state of the world that either can or can not be successfully contracted upon is a promising problem that merits further study. It is important because it identifies the potential content of contracts.

[31] Two cautionary points, however, should be noted about the prospect of trade at the interim. First, statements about the possibility of trading across all interim states are problematic in the case of reversible contingencies for precisely the same reasons that an ex ante contract can not arrange gains from trade in this case. Second, unlike the fixed-price game above, a trader may not have a dominant strategy in an arbitrary game in a particular state (α_B^*, α_S^*); his choice at the interim would therefore depend upon the strategy of his opponent. The specification of such a strategy leads back through Theorem 1 toward the inefficiency result of Theorem 9 in the case of reversible contingencies.

8 Appendix: Proofs of Results

8.1 Contracts and Mechanisms

The proof of the following lemma is straightforward.

Lemma 10 *If the mechanism $\hat{f}$ and the contract f satisfy $f = \hat{f} \circ v$, then:*

1. *$\hat{f}$ is incentive compatible with respect to the revelation of valuations if and only if f is incentive compatible with respect to revelation of types;*

2. *$\hat{f}$ is interim individual rational for each valuation of each agent if and only if f is interim individually rational for each type of each agent.*

Theorem 11 *Suppose that:*

1. *C is convex and the objective $\Phi(c, v_1, v_2)$ is concave in c for each v_1 and v_2.*

2. *The functions h_j and t_j are affine[32] in c for each agent j that has private information (i.e., $A_j \neq \emptyset$).*

Then the following statements are true:

1. *For every incentive compatible and interim individually rational contract f, there exists an incentive compatible and interim individually rational mechanism $\hat{f}$ that ex ante weakly dominates f:*

$$E_{[\underline{v}_1, \overline{v}_1] \times [\underline{v}_2, \overline{v}_2]} \left[\Phi \left(\hat{f}(v), v \right) \right] \geq E_A \left[\Phi \left(f(\alpha), v(\alpha) \right) \right]. \qquad (38)$$

2. *As a consequence of statement 1, if a given mechanism $\hat{f}$ solves the optimal mechanism problem, then the induced contract $f = \hat{f} \circ v$ solves the optimal contract problem. Conversely, if a contract f solves the optimal contract problem, then the mechanism $\hat{f}$ defined in the proof of statement 1 solves the optimal mechanism problem.*

Proof. Define the contract g by averaging f over all states that determine the same valuations: for $\alpha^* \in A$,

$$g(\alpha^*) = E_A[f(\alpha) | v(\alpha) = v(\alpha^*)].$$

The mechanism $\hat{f}$ that is sought is defined as $\hat{f} \circ v = g$. Inequality (38) follows from the concavity of Φ together with an application of Jensen's Inequality:

$$\begin{aligned} E_{[\underline{v}_1, \overline{v}_1] \times [\underline{v}_2, \overline{v}_2]} \left[\Phi \left(\hat{f}(v), v \right) \right] &= E_A \left[\Phi \left(g(\alpha), v(\alpha) \right) \right] \\ &= E_A \left[\Phi \left(E_A \left[f(\alpha^*) | v(\alpha^*) = v(\alpha) \right], v(\alpha) \right) \right] \\ &\geq E_A \left[E_A \left[\Phi \left(f(\alpha^*), v(\alpha^*) \right) | v(\alpha^*) = v(\alpha) \right] \right] \\ &= E_A \left[\Phi \left(f(\alpha), v(\alpha) \right) \right]. \end{aligned}$$

[32] That is, $h_j(\beta c + (1-\beta)c) = \beta h_j(c) + (1-\beta) h_j(c)$ for all $c \in C$ and $\beta \in [0, 1]$, and similarly for t_j.

Applying Lemma 10, the mechanism $\hat{f}$ is shown to satisfy *IC* and *IIR* by showing that the contract g has these properties. For notational simplicity, we do this for $j = 1$. For $\alpha_1^* \in A_1$, define

$$\begin{aligned} H_1'(\alpha_1^*) &= E_{A_2}\left[h_1(g(\alpha))\,|\alpha_1 = \alpha_1^*\right], \text{ and} \\ T_1'(\alpha_1^*) &= E_{A_2}\left[t_1(g(\alpha))\,|\alpha_1 = \alpha_1^*\right]. \end{aligned}$$

The fact that h_1 is affine implies

$$\begin{aligned} E_{A_1}[H_1(\alpha_1)\,|v_1(\alpha_1) = v_1(\alpha_1^*)] &= E_{A_1}\left[E_{A_2}[h_1(f(\alpha))]\,|v_1(\alpha_1) = v_1(\alpha_1^*)\right] \\ &= E_{A_2}[E_{A_1}[h_1(f(\alpha))|v_1(\alpha_1) = v_1(\alpha_1^*)]] \\ &= E_{A_2}[E_A[h_1(f(\alpha))|v(\alpha) = v(\alpha^*)]] \\ &= E_{A_2}[h_1(E_A[f(\alpha)|v(\alpha) = v(\alpha^*)])] \\ &= E_{A_2}[h_1(g(\alpha^*))] = H_1'(\alpha_1^*). \end{aligned}$$

A similar argument shows that

$$E_{A_1}[T_1(\alpha_1)\,|v_1(\alpha_1) = v_1(\alpha_1^*)] = T_1'(\alpha_1^*).$$

These equalities are now applied to demonstrate that g satisfies *IC* and *IIR* for agent 1. Incentive compatibility of f implies

$$H_1(\alpha_1^*)v_1(\alpha_1^*) + T_1(\alpha_1^*) \geq H_1(\alpha_1)v_1(\alpha_1^*) + T_1(\alpha_1) \tag{39}$$

for all $\alpha_1, \alpha_1^* \in A_1$. Because (39) holds for all $\alpha_1 \in A_1$, it follows that for all $\alpha_1^{**} \in A_1$,

$$\begin{aligned} &H_1\left(\alpha_1^*\right)v_1(\alpha_1^*) + T_1(\alpha_1^*) \\ &\geq E_{A_1}\left[H_1\left(\alpha_1\right)v_1(\alpha_1^*) + T_1(\alpha_1)\,|v_1\left(\alpha_1\right) = v_1\left(\alpha_1^{**}\right)\right] \\ &= H_1'(\alpha_1^{**})v_1(\alpha_1^*) + T_1'(\alpha_1^{**}). \end{aligned} \tag{40}$$

Because (40) holds for all $\alpha_1^* \in A_1$, it follows that

$$\begin{aligned} &H_1'\left(\alpha_1^*\right)v_1(\alpha_1^*) + T_1'(\alpha_1^*) \\ &= E_{A_1}\left[H_1\left(\alpha_1\right)v_1(\alpha_1^*) + T_1(\alpha_1)\,|v_1\left(\alpha_1\right) = v_1\left(\alpha_1^*\right)\right] \\ &= E_{A_1}\left[H_1\left(\alpha_1\right)v_1(\alpha_1) + T_1(\alpha_1)\,|v_1\left(\alpha_1\right) = v_1\left(\alpha_1^*\right)\right] \\ &\geq E_{A_1}\left[H_1'\left(\alpha_1^{**}\right)v_1(\alpha_1) + T_1'(\alpha_1^{**})\,|v_1\left(\alpha_1\right) = v_1\left(\alpha_1^*\right)\right] \\ &= H_1'(\alpha_1^{**})v_1(\alpha_1^*) + T_1'(\alpha_1^{**}), \end{aligned}$$

and so g satisfies *IC*. Turning to *IIR*, we have

$$\begin{aligned} &H_1'(\alpha_1^*)v_1(\alpha_1^*) + T_1'(\alpha_1^*) \\ &= E_{A_1}[H_1(\alpha_1)v_1(\alpha_1^*) + T_1(\alpha_1)|v_1(\alpha_1) = v_1(\alpha_1^*)] \geq r. \end{aligned}$$

■

8.2 Limited Observability

Proof of Theorem 1. As noted before the statement of the theorem, the issue to be proven is that a strategy used by an agent constrained by limited observability is necessarily finite. The subscript j denoting an agent is omitted here to simplify the notation. The use of γ by the agent implies that there exists for every α' a number $n' \in \mathbb{N}$ such that $\gamma(\alpha')$ is determined by $\alpha'_{n'-}$, i.e., γ is constant on the set $\{\alpha'_{n'-}\} \times A_{n-1+}$ consisting of all types that begin with the same initial string $\alpha'_{n'-}$. For $t \in \mathbb{N}$, define $A(t)$ as the set of all states α' for which $\gamma(\alpha')$ is *not* determined by α'_{t-}. The sets $\{A(t) | t \in \mathbb{N}\}$ satisfy $A(t) \supset A(t+1)$ for all $t \in \mathbb{N}$. If $A(n) = \emptyset$ for some n, then γ is finite. We now suppose that no such n exists and derive a contradiction. The complement $A(t)^c$ of $A(t)$ in A consists of all states α' for which $\gamma(\alpha')$ is determined by α'_{t-}. This set satisfies

$$\left\{\alpha'_{t-}\right\} \times A_{t-1+} \subset A(t)^c \tag{41}$$

for every $\alpha' \in A(t)^c$. The set $\{\alpha'_{t-}\} \times A_{t-1+}$ is open in the product topology on $A = \{0,1\}^{\mathbb{N}}$ when each set $\{0,1\}$ in the product is assigned the discrete topology. $A(t)^c$ is therefore open and so $A(t)$ is closed. The Tychonoff Theorem implies that A is compact in the product topology, which implies that there exists $\alpha^* \in \bigcap_{t=1}^{\infty} A(t)$. The value $\gamma(\alpha^*)$ is not determined by α^*_{t-} for any $t \in \mathbb{N}$, which contradicts the assumption that γ is usable by an agent constrained with limited observability. ■

Proof of Theorem 2. Let (γ_1, γ_2) be a Bayesian-Nash equilibrium with limited observability in the game $(M_1 \times M_2, \eta)$ that implements f. Let $n \in \mathbb{N}$ be sufficiently large that each strategy $\gamma_j(\alpha_j)$ is determined by $\alpha_{j,n-}$. Because $f = \eta \circ (\gamma_1, \gamma_2)$, the contract f is finite. It is now shown by contradiction that f is incentive compatible. If not, then there exists $\alpha_j^*, \alpha_j^{**} \in A_j$ and $\varepsilon > 0$ such that

$$E_{A_i}\left[u_j(\alpha_j^*, f(\alpha_j^{**}, \alpha_i))\right] - E_{A_i}\left[u_j(\alpha_j^*, f(\alpha_j^*, \alpha_i)))\right] > \varepsilon > 0. \tag{42}$$

Continuity of $v_j(\cdot)$ implies the existence of $k \in \mathbb{N}$ such that

$$\left|u_j(\alpha_j'', c) - u_j(\alpha_j', c)\right| < \frac{\varepsilon}{2} \tag{43}$$

for all $c \in C$ and for all $\alpha_j', \alpha_j'' \in A$ such that $\alpha''_{j,k-} = \alpha'_{j,k-}$. For $t \geq k$, it follows that

$$\left|E_A\left[u_j\left(\alpha_j, f(\alpha_j^{**}, \alpha_i)\right) | \alpha_{j,t-} = \alpha^*_{j,t-}\right] - E_{A_i}\left[u_j(\alpha_j^*, f(\alpha_j^{**}, \alpha_i))\right]\right| < \frac{\varepsilon}{2}, \text{ and} \tag{44}$$

$$\left|E_A\left[u_j\left(\alpha_j, f(\alpha_j^{*}, \alpha_i)\right) | \alpha_{j,t-} = \alpha^*_{j,t-}\right] - E_{A_i}\left[u_j(\alpha_j^*, f(\alpha_j^{*}, \alpha_i))\right]\right| < \frac{\varepsilon}{2}. \tag{45}$$

Substitution into (42) implies

$$E_A\left[u_j\left(\alpha_j, f(\alpha_j^{**}, \alpha_i)\right)\left|\alpha_{j,t-} = \alpha_{j,t-}^*\right.\right] - E_A\left[u_j\left(\alpha_j, f(\alpha_j^{*}, \alpha_i)\right)\left|\alpha_{j,t-} = \alpha_{j,t-}^*\right.\right] > 0 \tag{46}$$

for all $t \geq k$. Replacing f with $\eta \circ (\gamma_1, \gamma_2)$ produces an inequality that contradicts the assumption that (γ_1, γ_2) is a Bayesian-Nash equilibrium with limited observability.

Conversely, suppose that f is incentive compatible and finite. Finiteness implies the existence of $n \in \mathbb{N}$ such that $f(\alpha_1, \alpha_2)$ depends only upon $\alpha_{1,n-}$ and $\alpha_{2,n-}$, which allows us to write

$$f(\alpha_1, \alpha_2) = f(\alpha_{1,n-}, \alpha_{2,n-}). \tag{47}$$

Consider the game in which $M_j = A_{j,n-}$ for $j = 1, 2$ and $\eta(m_1, m_2) = f(m_1, m_2)$. Let $\gamma_1(\alpha_1) = \alpha_{1,n-}$ and $\gamma_2(\alpha_2) = \alpha_{2,n-}$. Because $f = \eta \circ (\gamma_1, \gamma_2)$, it is sufficient to show that these finite strategies define a Bayesian-Nash equilibrium with limited observability. This is also proven by contradiction. Suppose that

$$E_A\left[u_j\left(\alpha_j, \eta\left(\gamma_j\left(\alpha_j^{**}\right), \gamma_i(\alpha_i)\right)\right)\left|\alpha_{j,t-} = \alpha_{j,t-}^*\right.\right] > E_A\left[u_j\left(\alpha_j, \eta\left(\gamma_j\left(\alpha_j^{*}\right), \gamma_i(\alpha_i)\right)\right)\left|\alpha_{j,t-} = \alpha_{j,t-}^*\right.\right] \tag{48}$$

for agent j, some $\alpha_j^*, \alpha_j^{**} \in A_j$ and some $t \geq n$. This is equivalent to

$$E_A\left[u_j\left(\alpha_j, f(\alpha_j^{**}, \alpha_i)\right)\left|\alpha_{j,t-} = \alpha_{j,t-}^*\right.\right] > E_A\left[u_j\left(\alpha_j, f(\alpha_j^{*}, \alpha_i)\right)\left|\alpha_{j,t-} = \alpha_{j,t-}^*\right.\right]. \tag{49}$$

Inequality (49) can hold only if

$$E_{A_i}\left[u_j\left(\alpha_j, f(\alpha_j^{**}, \alpha_i)\right)\right] > E_{A_i}\left[u_j\left(\alpha_j, f(\alpha_j^{*}, \alpha_i)\right)\right] \tag{50}$$

for some $\alpha_j \in A_j$ such that $\alpha_{j,t-} = \alpha_{j,t-}^*$. Equation (47) implies that $f(\alpha_j^*, \alpha_i) = f(\alpha_j, \alpha_i)$ and so (50) contradicts the incentive compatibility of f. ■

Analysis of Case 3 in Example 2. Let $S(\alpha_{n-})$ denote the sum

$$S(\alpha_{n-}) = \frac{1-\delta}{\delta} \sum_{t=1}^{n-1} \delta^t a_t$$

for any $\alpha_{n-} = (a_t)_{1 \leq t \leq n-1}$. The formula

$$v(\{\alpha_{n-}\} \times A_{n-1+}) = \left[S(\alpha_{n-}), S(\alpha_{n-}) + \delta^{n-1}\right], \tag{51}$$

which follows from properties of the geometric series along with the assumption that $\delta > 0.5$, will be needed below.

The argument is by contradiction. Given $x \in (0,1)$, let k be the smallest element of $\mathbb{N}$ such that $v(\alpha) \geq x$ or $v(\alpha) \leq x$ can be decided based upon α_{k+1-}

for any choice of $\alpha_{k+1-} \in A_{k+1-}$. Case 2 implies that $k > 1$. The minimality of k and formula (51) together imply that there exists an initial string α^*_{k-} such that

$$S\left(\alpha^*_{k-}\right) < x < S\left(\alpha^*_{k-}\right) + \delta^{k-1},$$

so that $v(\alpha) \leq x$ or $v(\alpha) \geq x$ cannot be decided with certainty based upon α^*_{k-}. Define $\overline{\alpha} = (\overline{a}_t)_{t\in\mathbb{N}}$ and $\underline{\alpha} = (\underline{a}_t)_{t\in\mathbb{N}}$ as follows:

$$\overline{a}_t = \begin{cases} a^*_t & \textit{if } t \leq k-1 \\ 1 & \textit{if } t \geq k \end{cases};$$

$$\underline{a}_t = \begin{cases} a^*_t & \textit{if } t \leq k-1 \\ 0 & \textit{if } t \geq k \end{cases}.$$

Because $v(\overline{\alpha}) = S\left(\alpha^*_{k-}\right) + \delta^{k-1} > x$, it must be the case that $S(\overline{\alpha}_{k+1-}) \geq x$ else $v(\alpha) \leq x$ or $v(\alpha) \geq x$ could not be decided based upon $\overline{\alpha}_{k+1-}$. Because $v(\underline{\alpha}) = S\left(\alpha^*_{k-}\right) < x$, similar reasoning using formula (51) in the case of $\alpha_{n-} = \underline{\alpha}_{k+1-}$ implies that $S(\underline{\alpha}_{k+1-}) + \delta^k \leq x$. Combining these inequalities produces

$$\frac{1-\delta}{\delta} \cdot \delta^k = S(\overline{\alpha}_{k+1-}) - S(\underline{\alpha}_{k+1-}) \geq x - \left(x - \delta^k\right) = \delta^k,$$

where the first equality follows directly from the definitions of $S(\cdot)$, $\overline{\alpha}$ and $\underline{\alpha}$. This inequality cannot hold because $(1-\delta)/\delta < 1$ for $\delta > 0.5$. ■

8.3 Reversibility and Strong Irreversibility

Proof of Theorem 4. The proof is by induction on the length $n-1$ of the initial string α_{n-}. The case of $n-1=0$ is obvious, given the assumption that $v(A) = [\underline{v}, \overline{v}]$. Assume that statements 1 and 2 hold for initial strings of length $n-1$. We first show that $D_{(\alpha,_{n-},0)}$ and $D_{(\alpha_{n-},1)}$ are closed intervals whose intersection consists of a single point. It is then shown that statement 2 holds for all pairs of distinct initial strings of length n.

We begin by noting that $D_{(\alpha_{n-},0)}$ and $D_{(\alpha_{n-},1)}$ are closed sets. The Tychonoff Theorem implies that A is compact in the product topology when each set $\{0,1\}$ is assigned the discrete topology. The sets $\{\alpha' | \alpha'_{n-} = \alpha_{n-}, a'_n = 0\}$ and $\{\alpha' | \alpha'_{n-} = \alpha_{n-}, a'_n = 1\}$ are closed subsets of A_j in this topology and are therefore also compact. As noted in section 2.2, continuity of $v(\cdot)$ in the sense of Definition 2 is equivalent to its continuity of $v(\cdot)$ relative to this topology. Continuity of $v(\cdot)$ in this second sense implies that $D_{(\alpha_{n-},0)}$ and $D_{(\alpha_{n-},1)}$ are compact and therefore closed.

Because $D_{(\alpha_{n-},1)} \cup D_{(\alpha_{n-},0)} = D_{\alpha_{n-}}$ and $D_{\alpha_{n-}}$ is a closed interval by the induction hypothesis, the connectedness of $D_{\alpha_{n-}}$ implies that $D_{(\alpha_{n-},1)} \cap D_{(\alpha_{n-},0)} \neq \emptyset$. Assume by way of contradiction that $D_{(\alpha_{n-},0)}$ and $D_{(\alpha_{n-},1)}$ either fail to be closed intervals or else they are closed intervals whose interiors overlap.

Either of these statements can be true only if there exists $y_1, y_2 \in D_{(\alpha_{n-},0)}$ and $z \in D_{(\alpha_{n-},1)} \setminus D_{(\alpha_{n-},0)}$ with $y_1 < z < y_2$. Let

$$\overline{y} = \sup\{y \in D_{(\alpha_{n-},0)} | y \leq z\} \quad \textit{and} \quad \underline{y} = \inf\{y \in D_{(\alpha_{n-},0)} | y \geq z\}.$$

The points $\underline{y}$, $\overline{y}$ exist and are elements of $D_{(\alpha_{n-},0)}$ because this set is compact, and $\underline{y} < \overline{y}$ because otherwise $z = \underline{y} = \overline{y} \in D_{(\alpha_{n-},0)}$. It is also the case that

$$\overline{y} = \inf\{y \in D_{(\alpha_{n-},1)} | y \geq \overline{y}\} \quad \textit{and} \quad \underline{y} = \sup\{y \in D_{(\alpha,_{n-},1)} | y \leq \underline{y}\},$$

and consequently $\underline{y}, \overline{y} \in D_{(\alpha_{n-},1)}$. For $k = 1, 2$, $\underline{\alpha}^k$ and $\overline{\alpha}^k$ therefore exist such that $\underline{\alpha}^k_{n+1-} = (\alpha_{n-}, 0)$, $\overline{\alpha}^k_{n+1-} = (\alpha_{n-}, 1)$, $v(\underline{\alpha}^1) = v(\overline{\alpha}^1) = \underline{y}$, and $v(\underline{\alpha}^2) = v(\overline{\alpha}^2) = \overline{y}$. This contradicts the assumption that every contingency is strongly irreversible, and so $D_{(\alpha_{n-},0)}$ and $D_{(\alpha_{n-},1)}$ satisfy the conclusion.

Turning to statement 2 for $\alpha_{n+1-} \neq \alpha'_{n+1-}$, either $\alpha_{n-} = \alpha'_{n-}$ or $\alpha_{n-} \neq \alpha'_{n-}$. We have just shown in the first case that $D_{\alpha_{n+1-}} \cap D_{\alpha'_{n+1-}}$ contains one element. If $\alpha_{n-} \neq \alpha'_{n-}$, then the induction hypothesis together with the fact that $D_{\alpha_{n+1-}} \subset D_{\alpha_{n-}}$ and $D_{\alpha'_{n+1-}} \subset D_{\alpha'_{n-}}$ imply that $D_{\alpha_{n+1-}} \cap D_{\alpha'_{n+1-}}$ contains at most one element. ■

8.4 The Principal-Agent Problem

Proof of Theorem 5. Statement 2 of the theorem is all that remains to be proven. Let $((e_m, t_m))_{m\in\mathbb{N}}$ be a sequence of incentive compatible, interim individually rational and finite contracts that demonstrates the recordability of (e, t). Statement 1 implies that each contract in the sequence can be written as $(e_m(\alpha_{r-}), t_m(\alpha_{r-}))$, reflecting the fact that it is determined by the initial string α_{r-}. The sequence $(e_m(\alpha_{r-}), t_m(\alpha_{r-}))_{m\in\mathbb{N}}$ lies within the compact set $[0, \overline{e}] \times [\underline{t}, \overline{t}]$ and thus has a convergent subsequence. Because there are a finite number of initial strings of length $r - 1$, it is possible to select a subsequence $((e^*_m, t^*_m))_{m\in\mathbb{N}}$ of $((e_m, t_m))_{m\in\mathbb{N}}$ that converges for each α_{r-}. Define (e^*, t^*) as the limit of this subsequence:

$$(e^*(\alpha), t^*(\alpha)) = \lim_{m\to\infty} (e^*_m(\alpha_{r-}), t^*_m(\alpha_{r-})).$$

The contract (e^*, t^*) is finite and it inherits incentive compatibility and interim individual rationality from the contracts in the sequence $((e^*_m, t^*_m))_{m\in\mathbb{N}}$. As for the principal's ex ante utility, the continuity of the production function h and the inequality that demonstrates recordability of (e, t) together imply the desired result:

$$\int h(e^*(\alpha)) - t^*(\alpha) d\pi(\alpha) = \sum_{\alpha^*_{r-} \in A_{r-}} [h(e^*(\alpha_{r-})) - t^*(\alpha_{r-})] \pi(\alpha^*_{r-})$$
$$= \lim_{m\to\infty} \sum_{\alpha^*_{r-} \in A_{r-}} [h(e^*_m(\alpha_{r-})) - t^*_m(\alpha_{r-})] \pi(\alpha^*_{r-})$$
$$= \lim_{m\to\infty} \int h(e^*_m(\alpha)) - t^*_m(\alpha) d\pi(\alpha)$$
$$\geq \int h(e(\alpha)) - t(\alpha) d\pi(\alpha).$$

■

Proof of Theorem 6. A standard argument shows that the problem of finding an optimal mechanism in the principal-agent problem is equivalent to

$$\max_{\hat{e}(v), U_1(v)} \int h(\hat{e}(v)) + v\hat{e}(v) - U_1(v) d\mu(v) \quad s.t.: \tag{52}$$

(i) $\hat{e}(\cdot)$ is nondecreasing;

(ii) $U_1(v) = U_1(\underline{v}) + \int_{\underline{v}}^{v} \hat{e}(s)ds$, for all $v \in [\underline{v}, \overline{v}]$;

(iii) $U_1(v) \geq r$ for all $v \in [\underline{v}, \overline{v}]$.

Because of quasilinearity of utility, constraints (i) and (ii) of (52) are equivalent to *IC*,[33] while constraint (iii) (given (i) and (ii)) is *IIR*. We show that a solution to problem (52) can be approximated by finite contracts.

Theorem 4 implies that the sets $D_{\alpha_{n-}}$ defined in the theorem are intervals $[x^n_i, x^n_{i+1}]$ for $1 \leq i \leq m_n$. The intervals have strictly positive probability with respect to μ because of assumption 4. That μ is nonatomic (assumption 3) implies $x^n_i < x^n_{i+1}$. Continuity of $v(\cdot)$ also implies that for every $\varepsilon > 0$ there exists $\overline{n} \in \mathbb{N}$ such that $|x^n_{i+1} - x^n_i| < \varepsilon$ for all $n \geq \overline{n}$ and $1 \leq i \leq m_n$. Define $\xi^n(v)$ as

$$\xi^n(v) = \sup \{x^n_i \,|x^n_i \leq v, 1 \leq i \leq m_n\},$$

so that $\xi^n(v)$ rounds v downward to the nearest x^n_i. Let $\hat{e}^n(v) = \hat{e}(\xi^n(v))$. Define the transfer function $\hat{t}^n(v)$ by the formula

$$\hat{t}^n(v) = \overline{r} + \int_{\underline{v}}^{v} \hat{e}^n(s)ds - v\hat{e}^n(v).$$

The transfer $\hat{t}^n(v)$ is defined so that the mechanism $(\hat{e}^n(v), \hat{t}^n(v))$ satisfies *IC* and *IIR*. The functions $\hat{e}^n(v)$ and $\hat{t}^n(v)$ are constant on each of the open intervals (x^n_i, x^n_{i+1}) for $1 \leq i \leq m_n$.

[33] See Proposition 23.D.2 in Mas-Colell, Whinston and Green (1995).

We now show that $(\hat{e}^n(v), \hat{t}^n(v))_{n\in\mathbb{N}}$ converges uniformly to $(\hat{e}(v), \hat{t}(v))$. Uniform convergence of $(\hat{e}^n)_{n\in\mathbb{N}}$ to $\hat{e}$ follows by construction because $\hat{e}$ is a continuous function on a compact set and because the length of the intervals $\left[x_i^n, x_{i+1}^n\right]$ converges uniformly to 0 as $n \to \infty$. Turning to $(\hat{t}^n(v))_{n\in\mathbb{N}}$, let $\varepsilon > 0$ be arbitrary. Note that $\overline{v} - 2\underline{v} > 0$ because $\underline{v} < \overline{v} < 0$. Choose $n \in \mathbb{N}$ such that $\|\hat{e}^n - \hat{e}\| < \varepsilon / (\overline{v} - 2\underline{v})$, where $\|\cdot\|$ denotes the sup norm. Then

$$\begin{aligned}
\left|\hat{t}^n(v) - \hat{t}(v)\right| &= \left|\int_{\underline{v}}^{v} \hat{e}^n(s) - \hat{e}(s)\, ds + v\hat{e}(v) - v\hat{e}^n(v)\right| \\
&\leq \int_{\underline{v}}^{v} \left|\hat{e}^n(s) - \hat{e}(s)\right| ds + |v| \cdot \left|\hat{e}^n(v) - \hat{e}(v)\right| \\
&< (\overline{v} - \underline{v}) \frac{\varepsilon}{\overline{v} - 2\underline{v}} + |\underline{v}| \frac{\varepsilon}{\overline{v} - 2\underline{v}} \\
&= (\overline{v} - 2\underline{v}) \frac{\varepsilon}{\overline{v} - 2\underline{v}} = \varepsilon.
\end{aligned}$$

The transfer $\hat{t}^n(v)$ thus converges uniformly to the optimal transfer $\hat{t}(v)$.

The sequence of contracts $((e^n, t^n))_{n\in\mathbb{N}}$ that demonstrates the recordability of the optimal contract $(\hat{e}(v(\alpha)), \hat{t}(v(\alpha)))$ is defined by setting $(e^n(\alpha), t^n(\alpha))$ equal to the value of $(\hat{e}^n(v), \hat{t}^n(v))$ in the interior of $D_{\alpha_{n-}}$. The contract $(e^n(\alpha), t^n(\alpha))$ is obviously finite. It may differ from the composition of the mechanism $(\hat{e}^n(v), \hat{t}^n(v))$ with the valuation mapping $v(\cdot)$ at states α such that $v(\alpha) = x_i$ for some $1 \leq i \leq m_n$, and so the incentive compatibility and interim individual rationality of $(e^n(\alpha), t^n(\alpha))$ does not follow immediately from Theorem 10. The incentive compatibility and individual rationality of $(\hat{e}^n(v), \hat{t}^n(v))$, however, together with the fact that $\hat{e}^n(v)$ and $\hat{t}^n(v)$ are constant on each interval (x_{i-1}, x_i) imply that

$$t_i(v') - x_i \cdot \hat{e}^n(v') = t_i(v'') - x_i \cdot \hat{e}^n(v'') \geq r \tag{53}$$

for $v' \in (x_{i-1}, x_i)$ and $v'' \in (x_i, x_{i+1})$. It follows that $(e^n(\alpha), t^n(\alpha))$ satisfies *IC* and *IIR*. Statement (53) together with the uniform convergence of $\left((\hat{e}^n(v), \hat{t}^n(v))\right)_{n\in\mathbb{N}}$ to $(\hat{e}(v), \hat{t}(v))$ imply that $((e^n(\alpha), t^n(\alpha)))_{n\in\mathbb{N}}$ converges uniformly to (e, t). ■

8.5 The Model of Bilateral Trade

Proof of Theorem 7. Let $((p_m, t_m))_{m\in\mathbb{N}}$ be a sequence of contracts that demonstrates the recordability of (p, t). The first step is to construct for each $m \in \mathbb{N}$ an *IC* and *IIR* contract (p_m^*, t_m^*) that is interim payoff equivalent to (p_m, t_m) and whose value at α is determined by $\alpha_{B,b-}$ and $\alpha_{S,s-}$. Define $(p_m^*(\alpha'), t_m^*(\alpha'))$ by averaging (p_m, t_m) over all states $\alpha = (\alpha_B, \alpha_S)$ such that $\alpha_{B,b-} = \alpha'_{B,b-}$ and $\alpha_{S,s-} = \alpha'_{S,s}$:

$$p_m^*(\alpha') = \int_{A_{B,b-1+} \times A_{S,s-1+}} p_m(\alpha) d\pi(\alpha \left| \alpha_{B,b-} = \alpha'_{B,b-}, \alpha_{S,s-} = \alpha'_{S,s} \right.),$$

and

$$t^*_m(\alpha') = \int_{A_{B,b-1+}\times A_{S,s-1+}} t_m(\alpha) d\pi(\alpha \left| \alpha_{B,b-} = \alpha'_{B,b-}, \alpha_{S,s-} = \alpha'_{S,s} \right.).$$

For any $\alpha'_B \in A_B$, Theorem 3 implies that $\int_{A_S} p_m(\alpha_B, \alpha_S) d\pi_S(\alpha_S)$ is constant over the set of all α_B such that $\alpha_{B,b-} = \alpha'_{B,b-}$. It follows that:

$$\int_{A_S} p_m(\alpha'_B, \alpha_S) d\pi_S(\alpha_S) =$$
$$\int_{A_{B,b-1+}} \int_{A_S} p_m(\alpha_B, \alpha_S) d\pi_S(\alpha_S) d\pi_B \left(\alpha_B \left| \alpha_{B,b-} = \alpha'_{B,b-} \right. \right)$$
$$= \int_{A_{S,s-}} \int_{A_{B,b-1+}\times A_{S,s-1+}} p_m(\alpha_B, \alpha'_S)$$
$$\cdot d\pi \left(\alpha \left| \alpha_{B,b-} = \alpha'_{B,b-}, \alpha'_{S,s} = \alpha_{S,s-} \right. \right) d\pi_S(\alpha_{S,s-})$$
$$= \int_{A_{S,s-}} p^*_m(\alpha'_B, \alpha_S) d\pi_S(\alpha_{S,s-})$$
$$= \int_{A_S} p^*_m(\alpha'_B, \alpha_S) d\pi_S(\alpha_S).$$

The second equality is a change in the order of integration, the third applies the definition of $p^*_m(\alpha'_B, \alpha_S)$, and the last is true because $p^*_m(\alpha'_B, \alpha_S)$ is determined by $\alpha_{S,s-}$. Similar arguments prove that for all $\alpha'_B \in A_B$ and $\alpha'_S \in A_S$,

$$\begin{aligned}
\int_{A_S} t_m(\alpha'_B, \alpha_S) d\pi_S(\alpha_S) &= \int_{A_S} t^*_m(\alpha'_B, \alpha_S) d\pi_S(\alpha_S), \\
\int_{A_B} p_m(\alpha_B, \alpha'_S) d\pi_B(\alpha_B) &= \int_{A_B} p^*_m(\alpha_B, \alpha'_S) d\pi_B(\alpha_B), \textit{ and} \\
\int_{A_B} t_m(\alpha_B, \alpha'_S) d\pi_B(\alpha_B) &= \int_{A_B} t^*_m(\alpha_B, \alpha'_S) d\pi_B(\alpha_B).
\end{aligned}$$

The constraints of *IC* and *IIR* thus follow for (p^*_m, t^*_m) from the corresponding properties of (p_m, t_m). The interim expected utility function of each trader and the ex ante expected gains from trade (p^*_m, t^*_m) are the same as in (p_m, t_m).

The value of (p^*_m, t^*_m) at α is determined by $\alpha_{B,b-}$ and $\alpha_{S,s-}$, and the set $A_{B,b-} \times A_{S,s-}$ of all pairs $(\alpha_{B,b-}, \alpha_{S,s-})$ of such initial strings is finite. As a probability, $p^*_m(\alpha) \in [0, 1]$. The boundedness of the sets $v_B(A_B)$ and $v_S(A_S)$, assumption (31) of section 6, and *IIR* together imply that $(t^*_m(\alpha_{B,b-}, \alpha_{S,s-}))_{m\in\mathbb{N}}$ is bounded. By taking a subsequence (if necessary), it can thus be assumed without loss of generality that $(p^*_m(\alpha), t^*_m(\alpha))_{m\in\mathbb{N}}$ converges for all $\alpha \in A$. The contract (p^*, t^*) defined by

$$(p^*(\alpha), t^*(\alpha)) = \lim_{m\to\infty} (p^*_m(\alpha), t^*_m(\alpha))$$

for $\alpha \in A$ is thus well-defined. The contract (p^*, t^*) inherits *IC* and *IIR* from the contracts in the sequence, and it clearly satisfies statement 1 of the theorem.

Let $p^*_m(\alpha'_{B,b-}, \alpha'_{S,s})$ denote the value of $p^*_m(\alpha)$ at any state α such that $\alpha_{B,b-} = \alpha'_{B,b-}$ and $\alpha_{S,s-} = \alpha'_{S,s-}$. To verify statement 2 in the theorem, Lebesgue's Convergence Theorem implies

$$\begin{aligned} &\int_A (v_B(\alpha_B) - v_S(\alpha_S))\, p^*(\alpha) d\pi(\alpha) \\ &= \lim_{m\to\infty} \int_{A_B \times A_S} (v_B(\alpha_B) - v_S(\alpha_S))\, p^*_m(\alpha) d\pi(\alpha) \\ &= \lim_{m\to\infty} \int_{A_B \times A_S} (v_B(\alpha_B) - v_S(\alpha_S))\, p_m(\alpha) d\pi(\alpha) \\ &\geq \int_A (v_B(\alpha_B) - v_S(\alpha_S))\, p(\alpha) d\pi(\alpha), \end{aligned}$$

where the last two lines are the recordability inequality. ■

Proof of Theorem 8. Theorem 4 implies that the sets $D_{\alpha_j,n-}$ defined in the theorem are intervals $\left[x^n_{j,i}, x^n_{j,i+1}\right]$ for $j = B, S$ and $1 \leq i \leq m_n$. Assumption (31) of section 6 along with assumption 3 of the theorem imply that each of these intervals has a nonempty interior. Define $\xi^n_B(v_B)$ and $\xi^n_S(v_S)$ as follows:

$$\begin{aligned} \xi^n_B(v_B) &= \sup\left\{x^n_{B,i} \left| x^n_{B,i} \leq v_B, 1 \leq i \leq m_n \right.\right\}, \\ \xi^n_S(v_S) &= \inf\left\{x^n_{S,i} \left| v_S \leq x^n_{S,i}, 1 \leq i \leq m_n \right.\right\}. \end{aligned}$$

The function $\xi^n_B(\cdot)$ rounds a buyer's valuation downward while $\xi^n_S(\cdot)$ rounds a seller's valuation upward, in each case to the nearest boundary of one of the intervals $D_{\alpha_B,n-}$ or $D_{\alpha_S,n-}$, respectively. Continuity of $v_B(\cdot)$ and $v_S(\cdot)$ implies that $\lim_{n\to\infty} \xi^n_B(v_B) = v_B$ and $\lim_{n\to\infty} \xi^n_S(v_S) = v_S$ for all $v_B \in [\underline{v}_B, \overline{v}_B]$ and $v_S \in [\underline{v}_S, \overline{v}_S]$.

Theorem 11 implies that an optimal contract has the form $\left(\hat{p}^*(v(\alpha)), \hat{t}^*(v(\alpha))\right)$, where p^* and v^* solve the optimal mechanism problem

$$\max_{(p,t)} \iint (v_B - v_S)\, p(v_B, v_S) d\mu_B d\mu_S \quad \text{s.t. IC and IIR.} \tag{54}$$

Given the regularity conditions 3 and 4, Theorem 2 of Myerson and Satterthwaite (1983) characterizes a constant $k^* \in [0,1]$ such that $\hat{p}^*(v_B, v_S)$ has the form

$$\hat{p}^*(v_B, v_S) = \begin{cases} 1 & \text{if } V_B(v_B, k^*) \geq V_S(v_S, k^*); \\ 0 & \text{otherwise.} \end{cases}$$

Define the probability function $\hat{p}_n(v_B, v_S)$ as

$$\hat{p}_n(v_B, v_S) = \begin{cases} 1 & \text{if } V_B(\xi^n_B(v_B), k^*) \geq V_S(\xi^n_S(v_S), k^*); \\ 0 & \text{otherwise.} \end{cases}$$

The sequence $(\hat{p}_n)_{n\in\mathbb{N}}$ converges pointwise to $\hat{p}^*$. A comparison of $\hat{p}_n(v_B, v_S)$ with $\hat{p}^*(v_B, v_S)$ shows that $\hat{p}_n(v_B, v_S)$ satisfies inequality (2) of Myerson and Satterthwaite (1983) because $\hat{p}^*(v_B, v_S)$ satisfies it. It also inherits from $\hat{p}^*(v_B, v_S)$ the monotonicity properties required by Theorem 1 in their paper because $\xi_B^n(\cdot)$ and $\xi_S^n(\cdot)$ are nondecreasing. Formula (6) of their paper thus defines a transfer function $\hat{t}_n(v_B, v_S)$ such that the revelation mechanism $(\hat{p}_n, \hat{t}_n)$ satisfies *IC* and *IIR*. Like $\hat{p}_n(v_B, v_S)$, $\hat{t}_n(v_B, v_S)$ is constant on the interior of each set of the form $D_{\alpha_{B,n-}} \times D_{\alpha_{S,n-}}$.

The sequence $(p_n(\alpha), t_n(\alpha))_{n\in\mathbb{N}}$ that demonstrates the recordability of the optimal contract $\left(\hat{p}^*(v(\alpha)), \hat{t}^*(v(\alpha))\right)$ is defined as follows: for $\alpha = (\alpha_B, \alpha_S) \in A$, $(p_n(\alpha), t_n(\alpha))$ equals the value of $\left(\hat{p}_n(v), \hat{t}_n(v)\right)$ in the interior of $D_{\alpha_{B,n-}} \times D_{\alpha_{S,n-}}$. As in the proof of Theorem 6, it is straightforward to show that $(p_n(\alpha), t_n(\alpha))$ satisfies *IC* and *IIR* because $\left(\hat{p}_n(v), \hat{t}_n(v)\right)$ has these properties. It is clear that

$$\lim_{n\to\infty} p_n(\alpha) = \lim_{n\to\infty} \hat{p}_n(v(\alpha)) = \hat{p}^*(v(\alpha))$$

except at those states $\alpha = (\alpha_B, \alpha_S)$ for which either $v_B(\alpha_B)$ or $v_S(\alpha_S)$ is an endpoint of one of the intervals $\left[x_{j,i}^n, x_{j,i+1}^n\right]$ for $j = B$ or S, respectively, some $n \in \mathbb{N}$ and $1 \le i \le m_n$. Assumption 3 implies that this set of states has π-measure zero.

We conclude the proof by showing that the recordability inequality holds:

$$\begin{aligned}
& \int (v_B(\alpha_B) - v_S(\alpha_S))\, \hat{p}^*(v_B(\alpha_B), v_S(\alpha_S))\, d\pi(\alpha) \\
= & \iint (v_B - v_S) \left(\lim_{m\to\infty} \hat{p}_n(v_B, v_S)\right) d\mu_B(v_B)\, d\mu_S(v_S) \\
\le & \liminf \iint (v_B - v_S)\, \hat{p}_n(v_B, v_S)\, d\mu_B(v_B)\, d\mu_S(v_S) \\
= & \liminf \int (v_B(\alpha_B) - v_S(\alpha_S))\, p_n(\alpha_B, \alpha_S)\, d\pi(\alpha).
\end{aligned}$$

Because $(v_B - v_S)\, \hat{p}_n(v_B, v_S) \ge -|\underline{v}_B - \overline{v}_S|$, Fatou's Lemma implies the inequality. The last equality follows because $p_n(\alpha) = \hat{p}_n(v(\alpha))$ for π-a.e. α. By selecting a subsequence of $(p_n, t_n)_{n\in\mathbb{N}}$, the "lim inf" in the last line can be replaced with "lim", which completes the proof of recordability. ■

Proof of Theorem 9. For the recordable contract (p, t), Theorem 7 implies the existence of an incentive compatible, interim individually rational contract (p^*, t^*) such that: (i) (p^*, t^*) is state independent and hence constant; (ii) the ex ante gains from trade in (p^*, t^*) are as least as large as in (p, t). Suppose $p^* > 0$. Interim individual rationality implies

$$v_S(\alpha_S) \le \frac{t^*}{p^*} \le v_B(\alpha_B)$$

for all $\alpha_B \in A_B$ and $\alpha_S \in A_S$. It follows that $v_S(\alpha_S) \le v_B(\alpha_B)$ for all $\alpha_B \in A_B$ and $\alpha_S \in A_S$, which contradicts assumption 2 in the theorem. The contradiction implies that $p^* = 0$ and so the ex ante expected gains from trade in (p^*, t^*) are zero. Property (ii) above together with interim individual rationality therefore imply that they are also zero in (p, t). ■

References

[1] Al-Najjar, N. (2000): "Complexity as a Barrier to Competitive Imitation," working paper, Northwestern University.

[2] Anderlini, L., and L. Felli (1994): "Incomplete Written Contracts: Undescribable States of Nature," *Quarterly Journal of Economics,* 109, 1085-1124.

[3] Anderlini, L., and L. Felli (1998): "Describability and Agency Problems," *European Economic Review*, 42, 35–59.

[4] Chatterjee, K., and W. Samuelson (1983): "Bargaining under Incomplete Information," *Operations Research,* 31, 835–851.

[5] Dow, J. (1991): "Search Decisions with Limited Memory," *Review of Economic Studies*, 58, 1–14.

[6] Fudenberg, D., and J. Tirole (1991): *Game Theory.* MIT Press: Cambridge, MA.

[7] Hagerty, K., and W. Rogerson (1985): "Robust Trading Mechanisms," *Journal of Economic Theory*, 42, 94-107.

[8] Harasanyi, J. (1967-68): "Games with Incomplete Information Played by Bayesian Players," parts I, II, and III, *Management Science,* 14, 159-182, 320-334, 486-502.

[9] Hölmstrom, B., and R. Myerson (1983): "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51,1799-1819.

[10] Kelley, J: *General Topology.* Von Nostrand: New York. Reprinted as Vol. 27 in the series, *Graduate Texts of Mathematics.* Springer-Verlag: New York.

[11] Mas-Colell, A., M. Whinston and J. Green (1995): *Microeconomic Theory.* Oxford University Press: New York.

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

[13] Meyer, M. (1991): "Learning from Coarse Information: Biased Contests and Career Profiles," *Review of Economic Studies*, 58, 15–41.

[14] Milgrom, P., and N. Stokey (1982), "Information, Trade and Common Knowledge," *Journal of Economic Theory*, 26, 17-27.

[15] Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economics Studies*, 38, 175–208.

[16] Myerson, R., and M. A. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trade," *Journal of Economic Theory,* 29, 265–281.

[17] Radner, R. (2000): "Costly and Bounded Rationality in Individual and Team Decision-making," *Industrial and Corporate Change*, 9, 623-658.

[18] Rubinstein, A. (1993): "On Price Recognition and Computational Complexity in a Monopolistic Model," *Journal of Political Economy*, 101, 473–484.

[19] Rubinstein, A. (1998): *Modeling Bounded Rationality.* MIT Press: Cambridge, MA.

[20] Segal, I. (1999): "Complexity and Renegotiation: A Foundation for Incomplete Contracts," *Review of Economic Studies,* 66, 57-82.