ECONSTOR
*Make Your Publications Visible.*

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Holmstrom, Bengt; Myerson, Roger B.

**Working Paper**

# Efficient and Durable Decision Rules with Incomplete Information

Discussion Paper, No. 495

**Provided in Cooperation with:**
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

*Suggested Citation:* Holmstrom, Bengt; Myerson, Roger B. (1981) : Efficient and Durable Decision Rules with Incomplete Information, Discussion Paper, No. 495, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
https://hdl.handle.net/10419/220855

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Discussion Paper No. 495

# EFFICIENT AND DURABLE DECISION RULES WITH INCOMPLETE INFORMATION

by

Bengt Holmström and Roger B. Myerson

September 1981
revised, September 1982

J.L. Kellogg Graduate School of Management
Northwestern University
Evanston, Illinois 60201

Abstract. We compare six concepts of efficiency for economies with incomplete information, depending on the stage at which individuals' welfare is evaluated and on whether incentive constraints are recognized. An example is shown in which an incentive-efficient decision rule may be unanimously rejected by the individuals in the economy. We define durable decision rules, which can resist such unanimous rejection, and show that efficient durable decision rules exist.

Acknowledgements. We would like to thank Robert Aumann, Vincent Crawford, David Kreps, Charles Wilson, and Robert Wilson for very helpful discussions. Partial support from the N.S.F. and the Center for Advanced Study in Managerial Economics and Decision Sciences at Northwestern University is gratefully acknowledged.

# EFFICIENT AND DURABLE DECISION RULES WITH INCOMPLETE INFORMATION

## 1. Introduction

The concept of Pareto-efficiency is central in economics. For economies with complete information, this concept is straightforward: an economic allocation or decision is efficient if and only if there is no other feasible allocation that makes some individuals better off without making other individuals worse off. In modern economic theory, this concept of efficiency is the most important criterion for evaluating the performance of economic systems.

We say that there is incomplete information in an economy if its individual members have different private information at the time when the basic decisions about production and allocation must be made. That is, there is incomplete information whenever some individuals have information, about their preferences or endowments, which is not known by other individuals. For a seminal discussion of incomplete information, see Harsanyi [1967-8].

In an economy with incomplete information, the concept of efficiency becomes more difficult to define, because several new issues arise. Our goal in this paper is to systematically survey these issues and to develop the more sophisticated conceptual structure needed to talk about efficiency in economies with incomplete information.

In an economy with incomplete information, we must distinguish between decisions and decision rules. In this paper we use the term decision to refer to any plan for producing or allocating goods that can be implemented in any

state of the individuals' information. Because the individuals in the economy already know their information when these economic decisions are made, the actual decision will depend on the state of the individuals' information. A decision rule or mechanism is any specification of how the economic decisions are determined as a function of the individuals' information.

A welfare economist or social planner who analyzes the Pareto-efficiency of an economic system must use the perspective of an outsider, so he cannot base his analysis on the individuals' private information. That is, to judge whether a particular form of market organization is efficient in an economy with incomplete information, an outside economist can only analyze the decision rule induced by the market form. This is because he cannot predict what decision or allocation will be ultimately reached without knowing the individuals' private information. Thus, the proper object for welfare analysis in an economy with incomplete information is the decision rule, rather than the actual decision or allocation ultimately chosen. Furthermore, any efficiency criterion for evaluating decision rules must be defined independently of the unknown state of individuals' private information.

A definition of Pareto efficiency in an economy with incomplete information must look something like this: "a decision rule is efficient iff no other feasible decision rule can be found that may make some individuals better off without ever making any other individuals worse off." However, buried in this seemingly straightforward definition, there are at least three phrases whose interpretations are ambiguous and controversial.

First, there are several possible notions of feasibility for decision rules, depending on whether we require that a feasible decision rule must satisfy conditions of incentive compatibility or not. We will define and compare the notions of classical feasibility and incentive-feasibility in

Section 3.

Second, the notions of "better off" and "worse off" are ambiguous. Since we are considering a world of uncertainty, an individual is presumably better off whenever his expected utility is increased. But should his expected utility be computed as a function of his own private information, or of the join of all individuals' information, or should it be computed ex ante conditional on no private information at all? Section 4 will be devoted to this question, on the timing of welfare analysis.

The third point of ambiguity is about who is to "find" the potentially better decision rule: an outside planner or the informed individuals in the economy. This distinction is not important with complete information, because a planner is presumed to know all the relevant information in the economy. But in an economy with incomplete information, the planner does not know the individuals' types.

To understand the significance of this issue, notice that the concept of Pareto-efficiency has served two distinct purposes, one normative and one positive, in the study of economies with complete information. On the normative side, Pareto-inefficiency has been the primary justification for economists' recommendations to change market structures. That is, whenever a market was inefficient, an outside planner could propose ways to unambiguously improve the individuals' welfare. On the positive side, an argument (sometimes known as Coase's Theorem) has been made that one should generally expect to observe economies achieving Pareto-efficiency, when costs of bargaining can be ignored. If the market allocation were inefficient, then some individual could propose a reallocation which would make him better off and which all other individuals would be willing to accept. For economies with incomplete information, these normative and positive concepts of

efficiency may no longer coincide. Individuals might be able to unanimously agree on a change of decision rule which an outside planner could not have identified as better for all. Thus, normative concepts of Pareto-efficiency may admit decision rules which do not satisfy the positive criteria of Coase's Theorem.

The latter part of this paper is devoted to developing a concept of positive Coase-efficiency or durability for economies with incomplete information. In Section 5, we show that, without communication it cannot be common knowledge that the individuals would unanimously approve a change from a decision rule that is incentive-efficient, as defined in Section 4. However, in Section 6 we show an example in which one incentive-efficient decision rule clearly would be changed by a unanimous agreement of the individuals, if they can communicate and renegotiate their decision rule when each individual knows his type. Durable decision rules, which are resistant to such renegotiation, are defined in Section 7 and are shown to exist in Section 8.

Our paper has some direct predecessors in the literature, which ought to be pointed out. An early paper addressing itself to the issue of efficiency with incomplete information is Wilson [1978].[1/] Wilson's concept of efficiency is defined without recognizing incentive constraints, which appears appropriate only when there is a complete set of tradeable contingent claims, including claims contingent on private information. The proper concept of efficiency to use when incentive constraints are present was first explored in independent work by Harris and Townsend [1981], Holmström [1977] (p. 114-123) and Myerson [1979]. In all three, the same definition of incentive-efficiency was proposed, with an emphasis on incentive-compatible decision rules or mechanisms as the object of the definition. The present paper is a direct

continuation on this work. First we provide a taxonomy for different normative efficiency concepts, next we give an alternative characterization of incentive-efficiency in terms of common knowledge, and finally we investigate related concepts of efficiency or durability for positive economics.

## 2. Formulation

We will be concerned with the following abstraction of an economy. There are n individuals indexed i=1,...,n. For each individual i, let $T_i$ denote the finite set of possible types or information states of individual i. Each type $t_i$ in $T_i$ completely specifies i's preferences and beliefs, incorporating all of i's private information. Let

$$T = T_1 \times \ldots \times T_n \quad \text{and} \quad T_{-i} = T_1 \times \ldots \times T_{i-1} \times T_{i+1} \times \ldots \times T_n.$$

We will refer to $t = (t_1,\ldots,t_n)$ in T as an information state of the economy.

Individual i's beliefs concerning the state t are given by a probability measure $p_i$ on T, where $p_i(t)$ represents the subjective probability that i would assign to state t before he learns his type. We will assume that these probability distributions have the same zeros, that is,

(2.1) if $p_i(t) = 0$ then $p_j(t) = 0$ $\forall j, \forall i, \forall t \in T$.

We let $p_i(t_{-i}|t_i)$ denote the conditional subjective probability that i would assign to the event that $t_{-i}$ is the vector of other individuals' types, if his own type were $t_i$. By Bayes theorem,[2/]

$$p_i(t_{-i}|t_i) = p_i(t)/(\sum_{\hat{t}_{-i} \in T_{-i}} p_i(\hat{t}_{-i},t_i)).$$

(We use here the notational convention that, when t, $t_{-i}$, and $t_i$ appear in the same formula, then $t_{-i}$ represents the vector of all components other than $t_i$ in $t = (t_1,\ldots,t_n)$; and $(\hat{t}_{-i},t_i)$ represents the vector in T in which the $i^{th}$

component is $t_i$ and all other components are as in $\hat{t}_{-i}$.)

The economic problem is to select a <u>decision</u> or <u>allocation</u> $d_o$ from among a set of feasible decisions $D_o$. We note that $D_o$ can be given very different interpretations depending on the economic context. For example, $d_o$ could be an allocation of public and private goods, or a vector of reward schemes for the individuals conditional on some later observations, or a full Arrow-Debreu allocation of state-contingent claims.[3/]

It is natural to allow the choice of $d_o$ to be made in a randomized fashion if gains can be achieved thereby. For this purpose, let D be the set of generalized probability distributions over choices in $D_o$. If $D_o$ is finite with cardinality m then D is a simplex in $\mathbb{R}^m$. Our proofs will be carried out under this assumption.

Given any state t in T, the preferences of individuals are given by vonNeumann-Morgenstern utility functions $u_i(\cdot,t):D\to\mathbb{R}$, for i=1,...n. These may be derived utility measures, in that any uncertainty not included in t has been cared for by taking expectations. Furthermore since d is a probability distribution, all $u_i(d,t)$ are linear in d. Since $t_i$ incorporates all of individual i's private information, the functions $u_i(\cdot,\cdot)$ are common knowledge.

To sum up, the economy is completely specified by a list

$$\Gamma = (n,D_o,T_1,\ldots,T_n,p_1,\ldots,p_n,u_1,\ldots,u_n),$$

where n is the number of individuals, $D_o$ is the set of feasible decisions, $T_i$ is the set of agent i's information states, $p_i$ is the probability measure describing individual i's beliefs about the information state and $u_i$ describes i's preferences. All these component structures of $\Gamma$ are assumed to be common knowledge among all the individuals. In addition each individual i knows his own actual type (in $T_i$) as his private information.

## 3. Classical feasibility and incentive feasibility

Our concern is with concepts of efficiency in the economy described above. Efficiency would be straightforward to define if there were no private information or alternatively if the entire information state were to become publicly known. The set D would describe what is technologically feasible and an element d in D would be called efficient if and only if no other d' in D could be found which would make at least one individual better off without making anyone else worse off. Note that the efficient set would be common knowledge.

With private information matters get generally more complicated. Ordinarily it will no longer suffice to make comparisons between decisions alone since welfare may be improved by agreeing on decision rules which are contingent on the information state t. This may increase both insurance and production opportunities. Thus, as suggested above, the object for a definition of efficiency ought to be a decision rule $\delta:T \to D$. Since D includes the possibility of randomization, there is no need to consider randomized decision rules.

To decide which decision rules ought to be considered efficient, we must first resolve the issue of feasibility. One approach, which we will call classical, is to assume that there are no incentive problems involved in eliciting the necessary information $t=(t_1,\ldots,t_n)$ from individuals, when implementing any decision rule that maps information states to decisions. We let $\Delta$ denote the classically feasible set of decision rules, so that

$$\Delta = \{\delta:T \to D\}.$$

In general, a decision rule in $\Delta$ may not really be implementable, if it depends on information that individuals hold privately and that they do not want to reveal. Thus, feasibility of a decision rule is actually subject to

the constraint that individuals must have the incentive to report their private information truthfully. We say that a decision rule $\delta$ is incentive feasible or incentive compatible (in the Bayesian sense) iff

$$(3.1) \qquad \sum_{t_{-i}\in T_{-i}} p_i(t_{-i}|t_i)\, u_i(\delta(t),t) \geq \sum_{t_{-i}\in T_{-i}} p_i(t_{-i}|t_i)\, u_i(\delta(t_{-i},\hat{t}_i),t),$$

$$\forall i,\ \forall t_i \in T_i,\ \forall \hat{t}_i \in T_i.$$

The set of inequalities in (3.1), all linear contraints of $\delta$, guarantee that it is in the interest of each individual to report his type honestly if the other individuals do so. In the language of game theory, (3.1) implies that honest reporting strategies form a Nash equilibrium with $\delta$. We let $\Delta^*$ denote the set of incentive feasible decision rules, so that

$$\Delta^* = \{\delta \in \Delta \mid \delta \text{ satisfies (3.1)}\}.$$

In general, $\Delta^*$ is a proper closed convex subset of $\Delta$. However, the two sets may coincide in some special cases, in particular, if claims contingent on t can be traded.

One may conceive of mechanisms for choosing a decision in D that are more elaborate than the decision rules described above. One could construct mechanisms in which each individual may report messages other than just a simple statement of his type, as his input into the decision-making process. However, for any Bayesian equilibrium of individuals' reporting strategies in any mechanism for selecting the decision in D, there exists an equivalent incentive-compatible decision rule in $\Delta^*$. This idea is well known and has been called the revelation principle.[4/] The essential idea is that, given any equilibrium of reporting strategies in any mechanism, we can implement an equivalent incentive-compatible decision rule as follows: first we ask all

individuals to (simultaneously and confidentially) reveal their types; then we compute what each individual would have reported in the given equilibrium strategies with this type; and then we choose the (randomized) decision that the original mechanism would have chosen with these reports. If any individual had any incentive to lie to us in this decision rule, then he would have had an incentive to lie to himself in the original mechanism, which would contradict the premise that the original reporting strategies formed a Bayesian Nash equilibrium. This argument assures us that there is no loss of generality if we consider only the incentive-compatible decision rules in $\Delta^*$ for purposes of studying efficiency.

## 4. Ex ante, interim, and ex post efficiency concepts

How an individual's welfare should be measured depends crucially on what information he possesses at the time. Therefore, the proper concept of efficiency depends on when decision rules come up for welfare evaluation. Three evaluation stages appear relevant: ex ante, before individuals have received any private information; interim, when each individual has received his private information $t_i$, but does not know the other's information; and ex post, when the information state $t$ is public knowledge. The corresponding ex ante, interim, and ex post evaluations of a decision rule $\delta$ by individual $i$ are given by:

$$(4.1) \qquad U_i(\delta) = \sum_t p_i(t)\, u_i(\delta(t),t),$$

$$(4.2) \qquad U_i(\delta|t_i) = \sum_{t_{-i}} p_i(t_{-i}|t_i)\, u_i(\delta(t),t),$$

(4.3) $\quad U_i(\delta|t) = u_i(\delta(t),t).$

These three evaluation measures lead to three different notions of domination. We say that a decision rule $\gamma$ ex ante dominates $\delta$ iff

(4.4) $\quad U_i(\gamma) \geq U_i(\delta) \quad \forall i \in \{1,\ldots,n\}$

with at least one strict inequality;

$\gamma$ interim dominates $\delta$ iff

(4.5) $\quad U_i(\gamma|t_i) \geq U_i(\delta|t_i) \quad \forall i, \forall t_i \in T_i,$

with at least one strict inequality;

and $\gamma$ ex post dominates $\delta$ iff

(4.6) $\quad U_i(\gamma|t) \geq U_i(\delta|t) \quad \forall i, \forall t \in T,$

with at least one strict inequality.

Thus, if every individual would have preferred $\gamma$ over $\delta$ before learning his type, then $\gamma$ ex ante dominates $\delta$. If every individual would surely prefer $\gamma$ over $\delta$ when he knows his own type, whatever his type might be, then $\gamma$ interim dominates $\delta$. If every individual would prefer $\gamma$ over $\delta$ after learning all individuals' types, in any information state, then $\gamma$ ex post dominates $\delta$.

Notice that in the interim and ex post cases, domination requires (weakly) increasing expected utility for all possible types, not just for those in the actual information state of the economy. This requirement is necessary because a welfare economist, as an outsider, could not apply any concept of domination that depended on the individuals' actual private information. Therefore, we must compare $n$ utility measures in the ex ante case, $\sum_i |T_i|$ utility measures in the interim case, and $n \cdot |T|$ utility measures in the ex post case.

Our three notions of domination and two notions of feasibility ($\Delta$ and $\Delta^*$) together generate six potential concepts of efficiency. We say that a decision rule $\delta$ is ex ante classically efficient (or interim

classically efficient, or ex post classically efficient) iff there is no other decision rule $\gamma$ in $\Delta$ that ex ante (or interim, or ex post, respectively) dominates $\delta$. We let $\Delta_A$, $\Delta_I$, and $\Delta_P$ denote the sets of decision rules in $\Delta$ that are respectively ex ante, interim, and ex post classically efficient. We say that a decision rule $\delta$ is ex ante incentive-efficient (or interim incentive-efficient, or ex post incentive-efficient) iff $\delta$ is in $\Delta^*$ is and there exists no other incentive-compatible decision rule $\gamma$ in $\Delta^*$ that ex ante (or interim, or ex post, respectively) dominates $\delta$. We let $\Delta_A^*$, $\Delta_I^*$, and $\Delta_P^*$ denote the sets of decision rules that are respectively ex ante, interim, and ex post incentive-efficient.

The various concepts of efficiency can equivalently be represented through measurability restrictions on individual weights in a social welfare function. Let $\lambda_i$ map T into $\mathbb{R}_+$, for i =1,...,n, and consider the decision rules that maximize the social welfare function

$$W(\delta) = \sum_{i=1}^{n} \sum_{t\in T} \lambda_i(t)\, p_i(t)\, u_i(\delta(t),t).$$

If the $\lambda_i(t)$ depend on t arbitrarily then we get ex post efficient decision rules; if each $\lambda_i(t)$ depends only on $t_i$ then we get interim efficient rules; and if the $\lambda_i(t)$ are constants independent of t then we get ex ante efficient rules. In each case, one maximizes $W(\delta)$ over either $\Delta$ or $\Delta^*$, depending on whether classical or incentive efficiency is considered.

It is easy to see that

$$(4.7) \qquad \Delta_A \subseteq \Delta_I \subseteq \Delta_P, \quad \text{and} \quad \Delta_A^* \subseteq \Delta_I^* \subseteq \Delta_P^*.$$

That is, ex ante efficiency implies interim efficiency, which implies ex post efficiency, for either notion of feasibility. These inclusions reflect decreasing insurance opportunities when more information is released before agreeing on $\delta$. Therefore, if the individuals have agreed on an efficient

decision rule prior to obtaining their private information, there cannot be any other decision rule that would surely be better for all after they receive their private information. In the context of an exchange economy, Milgrom and Stokey [1982] have called this conclusion the no-trade theorem. Their theorem follows from the inclusion $\Delta_A \subseteq \Delta_I$.

Also, because $\Delta^* \subseteq \Delta$, we get the following inclusions:

$$(4.8) \qquad \Delta_A \cap \Delta^* \subseteq \Delta_A^*, \qquad \Delta_I \cap \Delta^* \subseteq \Delta_I^*, \qquad \Delta_P \cap \Delta^* \subseteq \Delta_P^*.$$

That is, any classically efficient decision rule that is incentive compatible is also incentive-efficient, in the appropriate sense. As an example, if each individual's utility function depends only on his own type, then a dictatorship by one individual is both classically efficient and incentive compatible.

It can also happen that $\Delta_P \cap \Delta^*$ may be empty, in which case no classically efficient mechanisms (in any sense) are incentive compatible. For example, suppose that there are two individuals, each individual has two equally likely types, $T_1 = \{1a,1b\}$ and $T_2 = \{2a,2b\}$, and each individual's type is stochastically independent of the other's. Suppose that there are two possible decisions, A and B, and the two individuals' payoff $(u_1,u_2)$ depend on the decisions and types as follows.

| | t = (1a,2a) | t = (1a,2b) | t = (1b,2a) | t = (1b,2b) |
|---|---|---|---|---|
| d = A | 6,0 | 0,0 | 2,2 | 0,0 |
| d = B | 0,6 | 2,2 | 0,0 | 2,2 |

For ex post classical efficiency, we must have

$$\delta(1a,2b) = B, \quad \delta(1b,2a) = A, \quad \delta(1b,2b) = B.$$

If type 1a pretended to be 1b in such a decision rule, while 2 was honest, then the expected utility for 1a would be 4; so for incentive compatibility we must have $\delta(1a,2a) = A$, to keep 1a from reporting "1b". But then type 2a would get lower expected utility from being honest than from claiming to be 2b ($1 \leq 3$). So no classically efficient decision rule can be incentive compatible in this example. Rosenthal [1978] has given other examples to illustrate similar points.

The taxonomy for normative efficiency concepts developed above should be useful as a reference framework. All notions, with the exception of $\Delta_P^*$, have been used earlier in the literature. However, in our view, only three of these notions (one for each evaluation stage) are relevant: ex ante incentive-efficiency $\Delta_A^*$, interim incentive-efficiency $\Delta_I^*$, and ex post classical efficiency $\Delta_P$. If the entire information state t were to become publicly known before the decision in D is chosen, then there would be no incentive problems and $\Delta_P$ would be the right efficiency concept to use. If the decision rule must be selected when each individual knows only his own type, then the incentive constraints (3.1) apply, and $\Delta_I^*$ is the right efficiency concept. If the decision rule can be selected before the individuals learn their types, but if the individuals cannot commit themselves ex ante to honestly report their types after they learn them, then $\Delta_A^*$ is the appropriate efficiency concept.[5/]

The distinction between situations in which interim or ex ante welfare analysis is relevant corresponds to the distinction between situations of <u>incomplete information</u> and <u>imperfect information</u> as it is sometimes made in the literature of game theory (see Harsanyi [1967-8]). That is, interim incentive-efficiency is the appropriate concept of efficiency for games with

incomplete information, in which the individuals already know their private information when the play of the game begins; and ex ante incentive-efficiency is the appropriate concept for games with imperfect information, in which the individuals learn their private information during the play of the game.

We will henceforth be concerned with situations in which the individuals select their decision rule at the interim stage. To simplify terminology, we may let incentive efficiency (unmodified) mean interim incentive-efficiency ($\Delta_I^*$) unless otherwise specified. The term ex post efficiency has always been used to mean what we have called ex post classical efficiency, and it seems reasonable to continue this usage.

## 5. Incentive efficiency and common knowledge.

If a decision rule $\delta$ is incentive-efficient (in the interim sense) then a social planner who does not know any individual's actual type could not propose any other incentive-compatible decision rule that every individual in any type is sure to prefer. However, there could possibly exist another incentive-compatible rule $\gamma$ and an information state t such that

$$(5.1) \qquad U_i(\gamma|t_i) > U_i(\delta|t_i) \quad \forall i.$$

In this case, if t were the actual information state, then all individuals would unanimously prefer $\gamma$ over $\delta$, each given his respective type.

Such unanimity is not effective for replacing $\delta$, however. The problem is that, even if (5.1) holds in state t, it may be that individual 1 would reverse his preference to favor $\delta$ if he learned that individual 2 also preferred $\gamma$ over $\delta$, since 2's preference would reveal new information to 1 about 2's type. If the individuals were to unanimously agree to change from $\delta$ to $\gamma$, then it would be common knowledge (in the sense of Aumann [1976]) that all individuals prefer $\gamma$ over $\delta$. (See Milgrom and Stokey [1982] and

Wilson [1978] for good discussions of this issue.) Thus, we should ask whether it could be common knowledge that all the individuals in the economy prefer $\gamma$ over $\delta$, when each individual knows only his own type.

We say that R is a common-knowledge event iff R is of the form $R = R_1 \times \ldots \times R_n$, where each $R_i \subseteq T_i$, and

$$(5.2) \qquad p_i(\hat{t}_{-i} | t_i) = 0, \qquad \forall t \in R, \ \forall \hat{t} \notin R, \ \forall i.$$

That is, if the information state of the economy is in the common-knowledge event R, then all individuals assign probability zero to the states outside of R. We say that $\gamma$ interim-dominates $\delta$ within R iff $R \neq \emptyset$ and

$$(5.3) \qquad U_i(\gamma | t_i) \geq U_i(\delta | t_i), \qquad \forall t \in R, \ \forall i,$$

with at least one strict inequality.

Theorem 1. An incentive-compatible decision rule $\delta$ is interim incentive-efficient if and only if there does not exist any common-knowledge event R such that $\delta$ is interim-dominated within R by another incentive-compatible decision rule.[6/]

Proof. The sufficiency part of this theorem is obvious (let R = T). For the necessary part, let $\delta$ be an incentive-efficient decision rule, and let $\gamma$ be an incentive-compatible decision rule. Suppose that $\gamma$ interim-dominates $\delta$ within R, and R is a common-knowledge event, contrary to the theorem. Let $\delta^*$ be

$$\delta^*(t) = \begin{cases} \gamma(t) & \text{if } t \in R, \\ \delta(t) & \text{if } t \notin R. \end{cases}$$

Then $$U_i(\delta^* | t_i) = \begin{cases} U_i(\gamma | t_i) & \text{if } t_i \in R_i, \\ U_i(\delta | t_i) & \text{if } t_i \notin R_i, \end{cases}$$

so $\delta^*$ interim dominates $\delta$ (globally). Furthermore $\delta^*$ is incentive compatible. To check this, consider first the case of an individual i whose type $t_i$ is in $R_i$. He could not gain in $\delta^*$ by claiming another type in $R_i$ because then $\delta^*$ coincides with $\gamma$, which is incentive compatible. Nor could he gain by claiming a type outside of $R_i$, where $\delta^*$ coincides with $\delta$, because $\delta$ is incentive compatible and $t_i$'s honest payoff in $\delta^*$ is at least as good as in $\delta$. On the other hand, if i's type t is not in $R_i$, then (by (2.1) and (5.2)) he is sure that no individuals are in their $R_j$ sets, and so he expects $\delta^*$ to coincide with the incentive-compatible rule $\delta$, whatever he reports, when the others are honest. Thus $\delta$ is interim-dominated (globally) by an incentive-compatible decision rule, which contradicts the assumption that $\delta$ is incentive-efficient. Q.E.D.

Theorem 1 implies that, if $\delta$ is incentive-efficient and each individual knows only his own type, then it cannot be common knowledge that the individuals unanimously prefer some other incentive-compatible decision rule over $\delta$. However, this result does <u>not</u> mean that the individuals could never reach a unanimous agreement to change from $\delta$ to some other incentive-compatible decision rule. It only means that if a unanimous agreement is reached then each individual must know more than just his own type; communication must have occurred.

Thus, one might now ask, if the decision rule for the economy is not imposed by an outside social planner, but instead is determined by individuals in the economy, after they have learned their types and in a situation of open communication, then what kinds of decision rules should we expect to observe? Would they necessarily be incentive-efficient? In the rest of this paper we shall attempt an introductory exploration of this complex issue.

## 6. Incentive-efficiency and durability: an example.

In an economy with complete information, Pareto-inefficient allocations are inherently unstable, precisely because the individuals could unanimously agree to some Pareto-superior allocation. Some economists, following Coase [1960] have therefore argued that we should expect to observe efficient allocations in any economy where there is complete information and bargaining costs are small. However, this positive aspect of efficiency does not extend to economies with incomplete information. To see why, let us consider a simple example.

Suppose that there are two individuals in the economy, and each individual may be one of two possible types. Individual 1 may be type 1a or 1b, individual 2 may be type 2a or 2b, and all four possible combinations of types are equally likely. There are three possible decisions called A, B, and C. The utility payoff of each individual from each decision depends only on his own type, as shown in the following table.

| | $u_{1a}$ | $u_{1b}$ | $u_{2a}$ | $u_{2b}$ |
|---|---|---|---|---|
| d = A | 2 | 0 | 2 | 2 |
| d = B | 1 | 4 | 1 | 1 |
| d = C | 0 | 9 | 0 | -8 |

In this example, individual 2 in either type and individual 1 in type 1a both prefer A over B and B over C. However if individual 1 is type 1b then his preference ordering is reversed and he strongly prefers C. Type 2b differs from 2a in that 2b has a greater aversion to decision C. (These are vonNeumann-Morgenstern utility numbers.)

Among all incentive-compatible decision rules, the following decision

rule $\delta$ uniquely maximizes the sum of the two individuals' ex ante expected utilities:

$$\delta(1a,2a) = A, \qquad \delta(1a,2b) = B,$$

$$\delta(1b,2a) = C, \qquad \delta(1b,2b) = B.$$

Notice that this decision rule selects decision C, type 1b's most preferred decision, if the types are 1b and 2a; but if 2's type is 2b (so that 2 is more strongly averse to C) then the decision rule selects B instead. To check that $\delta$ is incentive-compatible, notice that type 2a can get decisions A or C with equal probability if he is honest, or he can get B for sure if he lies and reports his type as 2b. Since both of these prospects give the same expected utility to 2a, he is willing to report his type honestly when $\delta$ is implemented.

This decision rule $\delta$ is incentive-efficient (in both the interim and ex ante senses), so no outsider could suggest any other incentive-compatible decision rule that makes some types better off without making any other types worse off than in $\delta$. But if individual 1 knows that his type actually is 1a, then he knows that he and individual 2 both prefer decision A over this decision rule $\delta$. Thus, rather than let $\delta$ be implemented, individual 1 in type 1a would suggest that decision A be implemented instead, and individual 2 would accept this suggestion.

Thus, although $\delta$ is an incentive-efficient decision rule, it is possible for the individuals to unanimously approve a change to some other decision rule (namely A-for-sure). Of course, this unanimity in favor of A over $\delta$ depends on 1's type being 1a, but consider what would happen if 1 were to insist on using $\delta$ rather than A. Individual 2 would infer that 1's type must be 1b. Then decision rule $\delta$ would no longer be incentive compatible, because both types of individual 2 would report "2b", to get decision B rather

than C.

Thus, if the individuals can redesign their decision rule when they already know their own types, then the decision rule $\delta$ could not be implemented in this example, even though it is incentive-compatible and incentive-efficient.[7/] In the terminology of the next section, we may say that this decision rule $\delta$ is incentive-efficient but not durable. Our next task is to develop a formal definition of durability that extends the positive sense of Pareto-efficiency to economies with incomplete information.

## 7. Durable decision rules

The essential idea to be developed in this section is that an incentive-compatible decision rule $\delta$ should be considered durable iff the individuals in the economy would never unanimously approve a change from $\delta$ to any other decision rule. Our problem is to formulate this idea rigorously.

Let us assume that $\delta:T \to D$ is an incentive-compatible decision rule, and that $\gamma$ is some alternative decision rule or mechanism being considered by the individuals. We do not assume that $\gamma$ is a direct-revelation mechanism, but that $\gamma$ can be any function of the form $\gamma:S_1\times\ldots\times S_n \to D$, where each $S_i$ is a nonempty finite set. The set $S_i$ represents the set of possible reports that individual i can select from, to communicate his informational input into the decision rule $\gamma$. We write $S = S_1\times\ldots\times S_n$. Of course, this notation allows for the possibility that $\gamma$ is a direct decision rule, in which case each $S_i = T_i$.

We want to establish whether the individuals in the economy would ever unanimously approve a change from the decision rule $\delta$ to $\gamma$. Thus, let us consider a voting game in which the alternative mechanism $\gamma$ will be implemented only if the individuals vote unanimously for $\gamma$. If any individual

votes against $\gamma$ then the status-quo decision rule $\delta$ will be implemented instead. We assume that the individuals vote simultaneously, and that only the outcome of the vote ($\delta$ or $\gamma$) will be learned by the individuals. (They cannot see each others' votes.) In such a voting game, we want to know whether there exists an equilibrium of voting strategies such that $\gamma$ is always rejected by at least one individual.[8/]

An individual's optimal voting strategy in this voting game should depend on how he would expect the mechanism $\gamma$ to be played if it were chosen, and what he would believe about the other individuals' types if they unanimously chose $\gamma$ over $\delta$. Thus, we shall need the following notation. For any individual $i$ and any type $t_i$, we let $r_i(t_i)$ denote the probability that $i$ would vote for $\gamma$ instead of $\delta$ if his type were $t_i$. For any $s_i$ in $S_i$, we let $\sigma_i(s_i|t_i)$ denote the probability that $i$ would use the report $s_i$ when $\gamma$ is implemented, if $t_i$ were his type and $\gamma$ won the vote. For any $t_{-i}$ in $T_{-i}$, we let $q_i(t_{-i}|t_i)$ denote the conditional probability that individual $i$ with type $t_i$ would assign to the event that $t_{-i}$ is the vector of other individuals' types, if he knew that they had all voted for $\gamma$ over $\delta$.

From these definitions we know that the quantities $(r,\sigma,q)$ must be nonnegative and must satisfy

$$(7.1)\qquad \sum_{t_{-i}} q_i(t_{-i}|t_i) = 1, \qquad \sum_{s_i \in S_i} \sigma_i(s_i|t_i) = 1, \qquad r_i(t_i) \leq 1, \qquad \forall i,\ \forall t_i \in T_i.$$

Given any types-vector $t = (t_1,\ldots,t_n)$ in $T$, and given any possible reports-vector $s = (s_1,\ldots,s_n)$ in $S$, we let

$$\sigma(s|t) = \prod_{j=1}^{n} \sigma_j(s_j|t_j).$$

Thus, $\sigma(s|t)$ is the probability that the individuals would give reports

$(s_1,\ldots,s_n)$ as input to $\gamma$, if their types were $(t_1,\ldots,t_n)$ and if $\gamma$ were implemented.

To show that $\delta$ is durable, we want to show that there is a Nash equilibrium of this voting game in which the alternative $\gamma$ is always rejected by at least one individual and $\delta$ is played honestly. The alternative $\gamma$ is always rejected iff

$$(7.2) \qquad \prod_{j=1}^{n} r_j(t_j) = 0, \quad \forall t \in T.$$

(This is equivalent to saying that there is at least one individual who rejects $\gamma$ in all information states.) If (7.2) holds, then honest behavior in $\delta$, together with the voting strategies $r = (r_1, \ldots, r_n)$ and reporting strategies $\sigma = (\sigma_1, \ldots, \sigma_n)$ in $\gamma$, form a Nash equilibrium of the voting game iff

$$(7.3) \qquad \sum_{t_{-i}} p_i(t_{-i}|t_i)\, u_i(\delta(t),t)$$

$$\geq \sum_{t_{-i}} p_i(t_{-i}|t_i)\Big(r_{-i}(t_{-i}) \sum_{s \in S} \sigma(s|t)\, u_i(\gamma(s_{-i},\hat{s}_i),t)$$

$$+ (1-r_{-i}(t_{-i}))\, u_i(\delta(t_{-i},\hat{t}_i),t)\Big)$$

$$\forall i, \quad \forall t_i \in T_i, \quad \forall \hat{s}_i \in S_i, \quad \forall \hat{t}_i \in T_i,$$

where

$$r_{-i}(t_{-i}) = \prod_{j \neq i} r_j(t_j).$$

That is, (7.3) asserts that individual i cannot gain by supporting the alternative $\gamma$ (and then reporting $\hat{t}_i$ if $\delta$ is implemented and

reporting $\hat{s}_i$ if $\gamma$ is implemented) when $t_i$ is his true type and all other individuals are expected to use their r and $\sigma$ strategies (for voting and reporting in $\gamma$) together with honest reporting in $\delta$. Since $\delta$ is incentive compatible, we know that no individual can expect to gain by rejecting $\gamma$ and then lying in $\delta$.

There is always a trivial Nash equilibrium in which $r_i(t_i) = 0$ for all i and $t_i$. This is an equilibrium because, as long as the other individuals are expected to vote against $\gamma$, each individual must expect that his vote cannot make any difference ($\delta$ will be implemented in any case); so he might as well vote against $\gamma$ too, even if he really would prefer $\gamma$. We must refine our analysis to exclude such equilibria, or else we would get the extreme result that every decision rule would be durable. (Recall the example in the preceding section). Thus, we must impose some kind of perfectness restrictions on the equilibria of the voting game (in the sense of Selten [1975]).

Let $E_i$ denote the event that all individuals other than i vote for the alternative $\gamma$ rather than the status quo $\delta$. Notice that i's vote only matters if $E_i$ occurs, since anyone can veto $\gamma$. Even if $E_i$ has zero probability, individual i would still have some posterior subjective probability distribution over $T_{-i}$ if $E_i$ occured, and we let $q_i$ denote this distribution. To exclude the trival equilibrium, we require that if, conditional on the event $E_i$, individual i with type $t_i$ would get higher expected utility in $\gamma$ than in $\delta$, then individual i with type $t_i$ must vote for $\gamma$. That is, for any type $t_i$ of any individual i,

$$(7.4) \qquad \text{if } \sum_{t_{-i}} q_i(t_{-i}|t_i)\, u_i(\delta(t),t) < \sum_{t_{-i}} \sum_{s\in S} q_i(t_{-i}|t_i)\, \sigma(s|t)\, u_i(\gamma(s),t)$$

$$\text{then } r_i(t_i) = 1.$$

We must also impose some restrictions on the reporting strategies $\sigma_i$ and the posteriors $q_i$ when $\gamma$ is chosen. Here we follow the basic ideas of sequential equilibria outlined by Kreps and Wilson [1982] and Selten [1975]. The reporting strategies $(\sigma_1,\ldots,\sigma_n)$ which the individuals would use in $\gamma$ must form a Nash equilibrium in the subgame when $\gamma$ is chosen, given the posterior beliefs $(q_1,\ldots,q_n)$. This condition can be expressed formally as follows:

$$(7.5)\qquad \sum_{t_{-i}} \sum_{s\in S} q_i(t_{-i}|t_i)\,\sigma(s|t)\,u_i(\gamma(s),t) \geq \sum_{t_{-i}} \sum_{s\in S} q_i(t_{-i}|t_i)\,\sigma(s|t)\,u_i(\gamma(s_{-i},\hat{s}_i),t),$$

$$\forall i,\ \forall t_i\in T_i,\ \forall \hat{s}_i\in S_i.$$

Condition (7.5) asserts that individual i with type $t_i$ should not expect any report $\hat{s}_i$ to be better for him in $\gamma$ than the report selected by his strategy $\sigma_i$.

Individual i should use Bayes theorem to compute his posterior $Q_i$ if the event $E_i$ occurs, so that

$$q_i(t_{-i}|t_i) = \frac{p_i(t_{-i}|t_i)r_{-i}(t_{-i})}{\sum_{\hat{t}_{-i}\in T_{-i}} p_i(\hat{t}_{-i}|t_i)\,r_{-i}(\hat{t}_{-i})}.$$

Unfortunately, this formula is not well-defined if some individual j different from i is expected to always reject $\gamma$ (so that $r_j \equiv 0$). In such a case, if i learned that the other individuals did unanimously approve $\gamma$ (event $E_i$) then he would infer that some mistake must have altered individual j's voting behavior. Then the "trembling hand" model of Selten [1975] and Kreps and Wilson [1982] gives us a way to characterize the rational posterior

distributions conditional on $\gamma$ being unanimously approved. Specifically, the posteriors $(q_1,\ldots,q_n)$ should satisfy the following condition

(7.6) there exists a sequence of voting strategies $\{r_1^k,\ldots,r_n^k\}_{k=1}^{\infty}$ such that:

$$r_j^k(t_j) > 0 \qquad \forall k,\ \forall j,\ \forall t_j \varepsilon T_j;$$

$$r_j(t_j) = \lim_{k\to\infty} r_j^k(t_j) \qquad \forall j,\ \forall t_j \varepsilon T_j;\ \text{and}$$

$$q_i(t_{-i}|t_i) = \lim_{k\to\infty} \frac{p_i(t_{-i}|t_i)\, r_{-i}^k(t_{-i})}{\sum_{\hat{t}_{-i}\varepsilon T_{-i}} p_i(\hat{t}_{-i}|t_i)\, r_{-i}^k(\hat{t}_{-i})} \qquad \forall i,\ \forall t_i \varepsilon T_i,\ \forall t_{-i} \varepsilon T_{-i},$$

where $r_{-i}^k(t_{-i}) = \prod_{j\neq i} r_j^k(t_j)$.

In the cases where $E_i$ has positive probability, (7.4) just reduces to Bayes theorem.

We say that $(r,\sigma,q)$ is an <u>equilibrium rejection</u> of $\gamma$, when the status quo is $\delta$, iff the conditions (7.1) through (7.6) are all satisfied. We say that $\delta$ <u>endures</u> $\gamma$ iff there exists some equilibrium rejection of $\gamma$, when the status quo is $\delta$. Finally, we say that $\delta$ is <u>durable</u> iff $\delta$ is an incentive-compatible decision rule and $\delta$ endures every alternative mechanism $\gamma:S\to D$, for every finite $S = S_1\times\ldots\times S_n$. In other words, if $\delta$ is durable then we can show rational voting equilibria in which the individuals never unanimously approve a change from $\delta$ to any other mechanism.

8. Existence of durable decision rules

In this section we will show that some decision rules actually are durable. To derive the results in this section, we will need to assume that no information state in T has zero probability, so that $p_i(t_{-i}|t_i) > 0$ for every individual i and every state t in T.

We say that an incentive-compatible decision rule $\delta$ is uniformly incentive-compatible iff

$$(8.1) \qquad u_i(\delta(t),t) \geq u_i(\delta(t_{-i},\hat{t}_i),t), \qquad \forall i,\ \forall t\in T,\ \forall \hat{t}_i\in T_i.$$

That is, $\delta$ is uniformly incentive-compatible if no individual would ever want to lie in $\delta$ about his type even if he knew the others' types, assuming that the others were planning to report their types honestly. For example, a decision rule that selects a constant decision in $D_o$ independently of t would be uniformly incentive-compatible. (If every individual's utility function is independent of the other individuals' types, then uniform incentive compatibility is equivalent to honesty being a dominant strategy for every individual.)

Theorem 2 If $\delta$ is uniformly incentive-compatible and interim incentive-efficient then $\delta$ is durable.

Proof Let $\gamma$ be any alternative mechanism, as in Section 7. Consider the voting game of Section 7, with the assumption that the individuals will report honestly in $\delta$ if it is implemented after the vote. As a function of his type, each individual must choose whether to vote for the alternative $\gamma$ or not, and what to report into the decision rule $\gamma$ if it is unanimously approved.

This is a finite game in extensive form, so it must have a perfect equilibrium in mixed strategies (see Selten [1975]). In the perfect equilibrium, let $r_i$ denote i's voting strategy and let $\sigma_i$ denote his reporting

strategy for $\gamma$, as before. The perfect equilibrium is the limit of a sequence of equilibria of perturbed games in which each individual always has some positive probability of voting for $\gamma$; so let $\{(r_1^k,\ldots,r_n^k)\}_{k=1}^{\infty}$ denote the voting strategies in this sequence of perturbed-game equilibria. We let $(q_1,\ldots,q_n)$ denote the limiting posteriors if $\gamma$ is chosen. These posteriors satisfy condition (7.6). The perfect equilibrium strategies also satisfy (7.4) and (7.5) for these limiting posteriors, since any perfect equilibrium is also a sequential equilibrium in the sense of Kreps and Wilson [1982]. Since (7.1) is trivially satisfied, only (7.2) and (7.3) remain to be shown.

This perfect equilibrium also satisfies one other condition that we did not require in Section 7:

$$(8.2)\quad \text{if } \sum_{t_{-i}} q_i(t_{-i}|t_i)\, u_i(\delta(t),t) > \sum_{t_{-i}} \sum_{s\in S} q_i(t_{-i}|t_i)\, \sigma(s|t)\, u_i(\gamma(s),t)$$

$$\text{then } r_i(t_i) = 0, \qquad \forall i,\ \forall t_i \in T_i.$$

That is, if i would prefer $\delta$ over $\gamma$ when the others all approved $\gamma$, then i must vote against $\gamma$.

The perfect equilibrium of the voting game is equivalent to a direct decision rule $\delta^*$, defined as follows:

$$\delta^*(t) = (1 - r(t))\, \delta(t) + r(t) \sum_{s\in S} \sigma(s|t)\, \gamma(s)$$

$$\text{where } r(t) = \prod_{j=1}^{n} r_j(t_j).$$

By (8.2), we know that no type of any individual could possibly do worse in $\delta^*$ than in $\delta$, since otherwise he would have rejected $\gamma$ and forced $\delta$. Thus

$$U_i(\delta^*|t_i) \geq U_i(\delta|t_i)$$

for all i and $t_i$.

Furthermore, $\delta^*$ is incentive compatible. To see this, notice that individual i could not gain by lying about his input into r or $\sigma$, in the formula for $\delta^*$, because $r_i$ and $\sigma_i$ were already chosen optimally for i in the perfect equilibrium. And individual i could not expect to gain by lying about his input into $\delta$ in the formula for $\delta^*$, because $\delta$ is uniformly incentive-compatible. That is, uniform incentive-compatibility implies that $\delta$ would still be incentive compatible given any information revealed by the fact that it has been chosen in the voting game. Thus $\delta^*$ is incentive compatible.

But $\delta$ was assumed to be incentive-efficient, so the preceding two paragraphs imply that $U_i(\delta^*|t_i) = U_i(\delta|t_i)$ for all i and $t_i$.

Suppose first that some type $t_1$ of individual 1 would expect to gain from $\gamma$ over $\delta$ conditional on the event $E_1$ (unanimity for $\gamma$). Then the probability of $E_1$ must be zero, or else $t_1$ would strictly prefer $\delta^*$ over $\delta$. So $r_{-1}(t_{-1})$ would be zero for all $t_{-1}$, since we have assumed that $p_1(t_{-1}|t_1) > 0$. Thus (7.2) would hold.

On the other hand, if no types of individual 1 would expect to gain from $\gamma$ over $\delta$, then we may assume that $r_1(t_1) = 0$ for all $t_1$; otherwise we could change to $r_1 \equiv 0$ (without changing Q or $\sigma$) and still satisfy (7.4)-(7.6). Then (7.2) would hold in this case as well.

Finally, we must check (7.3). By uniform incentive-compatibility, if (7.3) were ever violated then it would be violated for some $t_i$ with $\hat{t}_i = t_i$ in (7.3) and with positive probability of the event $E_i$. But in such a case, (7.6) would reduce to Bayes theorem; and (7.5) together with the interim domination of $\delta^*$ by $\delta$ would imply that (7.3) must hold.

Thus, the perfect equilibrium of the voting game gives an equilibrium rejection of $\gamma$. Q.E.D.

Suppose that individual 1 is the only individual with any private information, so that every other individual has only one possible type. Then there are no incentive constraints for the individuals other than 1 (they have nothing to report); and individual 1 already knows all the others' types when he reports his type into a decision rule. Thus every incentive-compatible decision rule is uniformly incentive-compatible in this case, and the following result (Holmström [1977]) is implied by Theorem 2.

Theorem 3. If there is only one individual with private information then every interim incentive-efficient decision rule is durable.

Myerson [1981] has studied situations in which there is one dictatorial individual, called the principal, who has the power to select the decision rule for the whole economy. In that paper, an incentive-compatible decision rule is called an expectational equilibrium iff every alternative mechanism has an equilibrium rejection in which only the principal casts the rejecting vote. Thus, expectational equilibria are durable. Also in that paper, the principal's neutral optimal mechanisms are defined, and it is shown that incentive-efficient neutral optima exist and are expectational equilibria. These results are significant for our current purposes because they imply our main existence theorem.

Theorem 4 There exists a nonempty set of decision rules that are both durable and incentive-efficient.

## 9. Concluding comments

The example in Section 6 shows that an incentive-efficient decision rule is not necessarily durable. To check that the decision rule $\delta$ in that example is not durable, let the alternative $\gamma$ select the outcome A in all states. Then there cannot be any equilibrium rejection of $\gamma$, because (7.4) implies that types 2b and 1a must vote for $\gamma$.

On the other hand, there can also exist decision rules that are durable but not incentive-efficient. For example, suppose that there are two individuals with two independent and equally likely types (1a, 1b; 2a, 2b), and there are two possible decisions, A and B. The two individuals get the same payoffs, as follows:

$$u_1(A,t) = u_2(A,t) = 2, \quad \forall t \in T;$$

$$u_1(B,t) = u_2(B,t) = \begin{cases} 3 \text{ if } t = (1a,2a) \text{ or } t = (1b,2b) \\ 0 \text{ if } t = (1a,2b) \text{ or } t = (1b,2a). \end{cases}$$

In this example, let $\delta(t) = A$ for all $t$. Then $\delta$ is not interim incentive-efficient but it is durable. The two individuals would both gain from changing to B when their types match; but in any voting game with any alternative mechanism, there is always an equilibrium rejection in which both individuals always use uninformative voting and reporting strategies (i.e., $r_i(t_i)$ and $\sigma_i(s_i|t_i)$ are independent of $t_i$). We implicitly assumed in Section 7 that the individuals would play noncooperatively in the voting game. Individuals cannot be forced to communicate effectively in a noncooperative game with incomplete information.

There is another point that requires some discussion and justification. In the proof of Theorem 2, we saw that a perfect equilibrium of the voting game necessarily satisfied an extra condition (8.2) that we did not require in Section 7. That is, we did not require that any type which would lose if the

alternative were approved must vote against the alternative. Thus, our equilibrium rejections are not necessarily perfect, as equilibria of the voting game.

One justification for omitting (8.2) in the definition of an equilibrium rejection is simply that condition (8.2) is not needed to eliminate the trivial equilibrium, in which everyone votes against the alternative because he does not expect his vote to matter. Alternatively, one could justify the asymmetry between including (7.4) and omitting (8.2) by imagining that there may be some infinitesimal cost to voting against the alternative, so that some individuals who would lose in the alternative might not bother to vote against it, since they are sure that someone else will veto it anyway.

Our practical reason for not requiring (8.2) was simply that we could not prove Theorems 4 and 5 if a durable decision rule's equilibrium rejections must also satisfy (8.2). To see what goes wrong, notice that condition (8.2) would become more restrictive if $u_i(\delta(t),t)$ were increased. In this way (8.2) is essentially different from the conditions (7.2) and (7.6), which would become less restrictive if the individuals' expected utility from honest participation in $\delta$ were increased. Nevertheless, it might be worth studying the properties of a revised definition of durability, in which equilibrium rejections are required to be fully perfect.

From another point of view, our definition of durability might seem too strong. We have required that a durable decision rule must have an equilibrium rejection against every alternative decision rule. However, the lack of an equilibrium rejection against some alternative would not necessarily undermine the status quo, unless one could argue that there is some individual who might be expected to actually propose this alternative. Otherwise, the voting game might never take place. Other definitions of

stability for decision rules might take this consideration into account in future research.

In this paper we have generally assumed that if the individuals in the economy unanimously commit themselves to a decision rule $\delta$ then they cannot recontract to alter the outcome selected by the decision rule afterwards. Without this assumption, the concept of stability for decision rules becomes even more complicated, since we must ask whether the individuals might unanimously vote to change the decision rule after they learn its outcome. For example, suppose that the decision rule $\delta$ was implemented and it selected the decision $d_o$ in $D_o$ as its outcome. Given this information, individual i with type $t_i$ would reassess his probability distribution over $T_{-i}$ to some $p_i'(\cdot|t_i)$, where Bayes theorem implies that, for all $t_{-i}$,

$$p_i'(t_{-i}|t_i)\left(\sum_{\hat{t}_{-i}\in T_{-i}} p_i(\hat{t}_{-i}|t_i)\,\delta(d_o|\hat{t}_{-i},t_i)\right) = p_i(t_{-i}|t_i)\,\delta(d_o|t).$$

If the decision $d_o$ (or, more precisely, the constant decision rule that always selects $d_o$) is durable in the context of the posterior economy

$$\Gamma' = (n,D_o,T_1,\ldots,T_n,p_1',\ldots,p_n',u_1,\ldots,u_n),$$

then any proposal to replace $d_o$ after $\delta$ selects it can always be vetoed. So $\delta$ can resist recontracting if it satisfies this posterior durability property for every possible outcome. However this property seems to be a very strong requirement to impose on a decision rule. We do not know what is the class of economies for which such decision rules may even exist in $\Delta^*$.

In any case, it is the authors' opinion that our definition of durability should be taken as only a first step in an attempt to build a theory of how individuals with private information might agree on a decision rule. We have found this to be a difficult problem, and we hope that others will join us in exploring it.

## Footnotes

1. For an extension of Wilson's approach to include economics with production, see Kobayashi [1980].
2. The prior probabilities $p_i(t)$ are hereafter only used in the definitions of ex ante domination and ex ante efficiency.
3. Harsanyi [1967-8] has argued that there is no need to make $D_o$ dependent on t. If some decisions are infeasible in a given state, then the individuals' preferences could be redefined so that they would all agree to avoid these decisions in this state.
4. This name was coined by Myerson, but the same principle was discovered by several others as well. Apparently the earliest recognition of the principle can be found in Gibbard [1973], while it has been most extensively explored in Dasgupta, Hammond, and Maskin [1979].
5. Wilson's [1978] concept of coarse efficiency corresponds to our set $\Delta_I$, and his fine efficiency corresponds to our $\Delta_P$. However, in Wilson's work, the decision set $D_o$ includes claims contingent on the information state, so that $\Delta_I$ and $\Delta_I^*$ are essentially equivalent in his case.
6. An analogous characterization of classical interim efficiency is found in Kobayashi [1980].
7. Our assumption that individuals can commit themselves to a decision rule at the interim stage (without possibility of recontracting ex post) might at first seem inconsistent with our assumption that the individuals cannot commit themselves to a decision rule ex ante, before they learn their types. But recall that, we are assuming that the individuals already have their private information about their preferences and endowments when they meet to make their economic plans and decisions. That is, we are studying

economies in which the ex ante stage, in which no individual has learned any private information, has already passed (if it ever indeed existed) so that "ex ante" commitments are impossible.

8. More complex voting games could be considered, but this one seems to be the simplest in which the question of unamimous rejection of $\delta$ can be considered. Furthermore, this voting game may be more general than it initially appears. For example, consider any multistage voting procedure, with many alternatives on the agenda, such that in the first stage any individual can anonymously terminate discussion and veto all alternatives to $\delta$. If we reinterpret the mechanism $\gamma$ as including (in normal form) all stages of this procedure after the first stage, then the first stage is equivalent to our simple voting game.

## References

Aumann, R. [1976],"Agreeing to Disagree," Annals of Statistics 4, 1236-1239.

Coase, R. H. [1960], "The Problem of Social Cost," Journal of Law and Economics 3, 1-44.

Dasgupta, P., P. Hammond, and E. Maskin [1979], "The Implementation of Social Choice Rules; Some General Results on Incentive Compatibility," Review of Economic Studies 46, 185-216.

Gibbard, A. [1973], "Manipulation of Voting Schemes: A General Result," Econometrica, 41, 587-602.

Harris, M. and R. M. Townsend [1981], "Resource Allocation Under Asymmetric Information," Econometrica 49, 33-64.

Harsanyi, J. C. [1967-8], "Games with Incomplete Information Played by Bayesian Players," Management Science 14, 159-182, 320-334, 481-502.

Holmström, B. [1977], "On Incentives and Control in Organizations," Ph.D. dissertation, Stanford University.

Kobayashi, T. [1980], "Equilibrium Contracts for Syndicates with Differential Information," Econometrica 48, 1635-1666.

Kreps, D. and R. Wilson [1982], "Sequential Equilibria," Econometrica 50, 863-894.

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

Myerson, R. B. [1979], "Incentive Compatibility and the Bargaining Problem," Econometrica 47, 61-74.

Myerson, R. B. [1981], "Mechanism Design by an Informed Principal," discussion paper, Center for Math Studies, Northwestern University. To appear in Econometrica.

Rosenthal, R. W. [1978], "Arbitration of Two-Party Disputes Under Uncertainty," Review of Economic Studies 45, 595-604.

Selten, R. [1975], "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," International Journal of Game Theory 4, 25-55.

Wilson, R. [1978], "Information, Efficiency, and the Core of an Economy," Econometrica 46, 807-816.