CALIFORNIA INSTITUTE OF TECHNOLOGY
Division of the Humanities and Social Sciences
Pasadena, California 91125

INCENTIVE COMPATIBILITY AND INCOMPLETE INFORMATION

John O. Ledyard*

Social Science Working Paper

Number 187

October 1977

---

*Both the National Science Foundation and the Fairchild Foundation are gratefully acknowledged for their support. Early parts of this work were completed at The Center for Mathematical Studies in Economics and Management Science, Northwestern University. I thank Ted Groves, Andy Postlewaite, and John Roberts for helpful conversations.

INCENTIVE COMPATIBILITY AND INCOMPLETE INFORMATION

by

John O. Ledyard

It is by now reasonably well known that when informationally decentralized processes are used to make collective choice decisions or to allocate resources, individuals may find it in their interest to distort the information they provide and that these distortions may lead to non-optimal group decisions. In the social choice context, this has been formalized in the Gibbard-Satterthwaite Theorem [2, 14] which states that all non-dictatorial rules will have this property. In a different context, Hurwicz [7] has shown that there is a private goods neo-classical exchange economy such that any decentralized mechanism which selects Pareto-optimal allocations and which has a no-trade option will have this property. Roberts [13] has provided a similar example in the public goods context. Other work (e.g., Green-Laffont [3], Groves-Loeb [4], Hurwicz [8] and Walker [15]) indicates that, for mechanisms designed to select efficient outcomes, in most environments some agent will have an incentive to misrepresent his information and thus to manipulate the mechanism. All these results lead one to the conjecture that it is almost impossible to design any mechanism for group decisions which is compatible with individual incentives and efficiency.

One possible way out of this dilemma can be identified by looking at the implicit informational assumptions behind the theorems concerning the impossibility of incentive compatibility. In particular, it is usually assumed that individuals have enough information, or can collect enough, to be able to compute an appropriate manipulative response. In general, this involves complete knowledge of the allocation mechanism and of the responses of the other individuals. One might conjecture that with less information individuals may be reluctant to misrepresent their input to the mechanism since they cannot be sure of achieving gains. Thus, some form of incentive compatibility might occur under more realistic informational assumptions. Countering this conjecture is the argument that finding a response which is better than the truthful response for an individual need not involve any information about the others at all. To see this, suppose that all consumers have chosen their responses but one is allowed to vary his while the others keep theirs fixed. If there is a better strategy for this individual, one would expect that he would eventually find it. If one accepts this view of strategy selection then the theorems above apply independently of any assumptions about initial information. However, we would argue that the individual who finds his manipulation in this fashion is getting something for free since he is never penalized for choices which are worse for him than the truth.

In this paper, we examine the impact on the incentive compatibility of a mechanism of two assumptions: (1) each individual possesses incomplete information and (2) each individual must select his representation before he can gather further information and once this strategy is selected, he may not revise it. Thus, a bad selection is penalized. These assumptions seem to be the most favorable conditions for attaining incentive compatibility of some mechanisms even though they are not incentive compatible under complete information. However, our conclusion will be that *the introduction of complete information alters none of the incentive properties of most mechanisms.* In particular, we will establish two general propositions. In section 2.B we show that an allocation mechanism (social choice function) is incentive compatible under incomplete information if and only if it is incentive compatible under complete information. In section 2C we show that a differentiable allocation mechanism will lack incentive compatibility in most environments under complete information only if it lacks incentive compatibility under incomplete information in most environments for most prior beliefs.

We begin in Section 1 by introducing the standard model of behavior under complete information. In Section 2A we extend the model to behavior with incomplete information. In Section 2B we characterize the situations in which mechanisms are incentive compatible under incomplete information. In Section 2C we characterize the situations in which the lack of incentive compatibility under incomplete information is generic. Section 3 contains some concluding remarks about the relationship of these results to others in the literature.

1. Complete Information

A. Abstract Allocation Mechanisms*

An allocation mechanism is often described formally by a communication process (a language and a collection of response rules) and an outcome rule describing how equilibrium messages get translated into allocations. (See, e.g., Hurwicz [6] or Reiter [12].) Since much of this usual description is unnecessary for the analysis in this paper, I will abstract entirely from the details and will formalize an allocation mechanism as a mapping from environments to allocations.

To be more precise an (economic) environment is $e = \{I, A, e^1, \ldots, e^n\}$ where I is the index set of n agents (indexed $i = 1, \ldots, n$), A is the space of feasible allocations and e^i is the characteristic of the *ith* agent.† Included in e^i is a description of i's preferences, $\geq (e^i)$, over the space of allocations. Throughout this paper, once I and A are chosen, they will no longer be allowed to vary. Thus one can, I hope without confusion, let E^i be the space of possible characteristics for agent i and let $E \equiv E^1 x \ldots x E^n$ be the space of all environments (given I and A). An environment is $e \in E$.

---

*We use the phrase "allocation mechanism" throughout this paper. However, one can just as easily read "social choice function" or "collective decision rule" without any difficulty.

†In a pure exchange environment, n is the number of consumers (agents), A is the set of joint net trades which add up to zero, and e^i includes a description of i's preferences.

Now given an environment, an allocation mechanism selects a feasible allocation, that is, an allocation mechanism is a function*, Φ, from E to A, where $\Phi(e) \in A$ is the allocation selected when the environment is e.

---

*Certain known mechanisms, such as the competitive process, may be set valued for some environments. However, I would argue that these are incompletely specified since ultimately only a single allocation can occur. Thus, a complete version of the competitive process would include the rule for selection of the allocation from the set of Walrasian allocations. If such a rule is a random device, then the complete mechanism selects gambles over the space of net trades, T. In this case, by letting A be the space of probability measures on T, M(T), one is back to our framework where the mechanism is a function from E to A = M(T). This method is used in [9].

B. <u>Incentive Compatibility</u>

The key notion behind incentive compatibility and the potential lack thereof, is the idea that the allocation mechanism cannot recognize the difference between an agent's true characteristic and his reported characteristic. In fact, the mechanism operates only to map reported characteristics into allocations. Thus, an agent, capable of taking full advantage of this fact, might be able to manipulate the mechanism, by reporting a false characteristic, and thereby leave himself better off than if he reported his true characteristic. A mechanism for which no agent can gain in this manner is called incentive compatible.

More formally, let $Z^i \equiv E^1 x \ldots x E^i \text{ x } E^{i+1} \text{ x} \ldots \text{x } E^n$ (the n-1 Cartesian product of the other agents' spaces of characteristics). We will use the vectors e and (e^i, z^i) interchangeably. No confusion should occur. We can now state

<u>Definition 1</u>: [Hurwicz] (a) An allocation mechanism $\Phi: E \to A$ is *incentive compatible for* i in $e \in E$ if and only if $\Phi(e^i, z^i) = \Phi(e) \geq (e^i) \; \Phi(s^i, z^i)$ for all $s^i \in E^i$. (That is, the outcome if i reports e^i is at least as preferred under his true preferences, $\geq (e^i)$, to any he could achieve through manipulation.)

(b) An allocation mechanism $\Phi: E \to A$ is *incentive compatible* on $B \subseteq E$ if and only if Φ is incentive compatible for i in e for all $e \in B$ and all i = 1, ..., n.

This definition is that introduced by Hurwicz [7] and is equivalent to non-manipulability of social choice functions.

2. Incomplete Information

A. Model and Definitions

The simplest way to model the incentive problem of an agent who lacks complete information and who must choose his representation, $s^i \in E^i$, without recourse is to assume that i has some prior beliefs about the parameters he does not know and cannot control. One especially compelling model of this type can be found in Harsayani [5]. The model we use is adapted from his. Other applications of his model to incentive issues can be found in [1] and [11].

We broaden the interpretation of an agent's characteristic, e^i, to include not only preferences and endowments but also beliefs about the environment and allocation mechanism. Thus, associated with each $e^i \in E^i$ is a probability measure on E representing i's prior beliefs.* More formally, we make the following assumption which will be maintained throughout this paper. Remember that $Z^i = E^1 x \ldots x\ E^{i-1}\ x\ E^{i+1}\ x \ldots x\ E^n$.

Assumption A.0: For each i = 1, ..., n,

(a) Z^i is endowed with a Hausdorff topology. $M(Z^i)$ is the space of all probability measures on $(Z^i, B(Z^i))$ where $B(Z^i)$ is the set of Borel subsets of Z^i. $M(Z^i)$ is endowed with the topology of weak convergence. The above also applies to A and M(A).

(b) With each $e^i \in E^i$ is associated a bounded, measurable (Von Neumann-Morgenstern) utility function, $u(\cdot, e^i)$, such that if μ, $\mu^1 \in M(A)$ are two random gambles over allocations in A then i prefers μ^1 to μ if

*As in Harsayani, these priors can be beliefs about what others believe about others' beliefs about etc.

and only if $\int u^i(a,e^i)\ d\mu^1 > \int u^i(a,e^i)\ d\mu$.

(c) With each $e^i \in E^i$ is associated a unique element $\eta \in M(Z^i)$ representing i's beliefs. We let the function $\psi^i : E^i \to M(Z^i)$ summarize this association. The space Ψ^i of all such functions is endowed with the product topology.

(d) We consider only those allocation mechanisms $\Phi: E \to A$ which are measurable on Z^i for all i.

Using this assumption we can now broaden the definition of incentive compatibility to cover situations with incomplete information.

Definition 2: (a) An allocation mechanism is said to be incentive compatible in e given beliefs ψ, where $\psi = (\psi^1, \ldots, \psi^n)$, if and only if, for i = 1, ..., n, $\int u^i\ [\Phi(e^i,z^i),e^i]\ d\psi^i\ (e^i) \geq$

$$\int u^i\ [\Phi(s^i,z^i),e^i]\ d\psi^i(e^i) \text{ for all } s^i \in E^i.$$

(That is, no i can have an expected utility from reporting truthfully which is lower than that achieved through some misrepresentation.)

(b) An allocation mechanism is said to be incentive compatible on $B \subseteq E \times \Psi$ (where $\Psi \equiv \Psi^1 \times \ldots \times \Psi^n$) if and only if Φ is incentive compatible in e given ψ for all $(e,\psi) \in B$.

Remark: Definition 2 represents a true broadening of Definition 1 since $M(Z^i)$ includes measures with mass concentrated on single points. Let $\eta^i_{z^i} \in M(Z^i)$ be the measure such that $\eta^i_{z^i}(\{z^i\}) = 1$. Let $\overline{\psi}^i(e^i) = \eta^i_{z^i}$. Then Φ is incentive compatible in (e^i,z^i) if and only if Φ is incentive compatible in e given $\overline{\psi}^i$. Thus, incentive compatibility on $E \times \Psi$ implies incentive compatibility on E. Note, however, that incentive compatibility in e given $\overline{\psi}$ for all $e \varepsilon E$ does not necessarily imply incentive compatibility in E.

Remark: One can easily argue that the model of incomplete information does not allow for enough uncertainties which i may face. For example, it is possible that i may not know how reported characteristics are transformed into allocations. That is, he may not understand completely how the allocation mechanism works. This additional uncertainty can be modeled by introducing another random variable, x, beliefs about x, $\gamma^i \varepsilon M(X)$, and a mapping $\Phi: E \times X \to A$. Then i's expected utility from reporting s^i, if e^i is his true characteristic, is

$$\int\int u^i(\Phi(s^i,z^i,x),\ e^i)\ d\gamma^i\ d\psi^i\ (e^i).$$

Let $v^i(s^i,z^i,e^i) \equiv u^i(\Phi(s^i,z^i,x),\ e^i)\ d\gamma^i$ and let

$\int v^i(s^i,z^i,e^i)\ d\psi^i(e^i)$ replace $\int u^i(\Phi(s^i,z^i),\ e^i)\ d\psi^i(e^i)$ in definition 2. None of the results reported in this paper would be altered by this procedure. Thus we will ignore uncertainty about the allocation mechanism.

Remark: Returning to section 1 for a moment, one notes that the definition of incentive compatibility (definition 1) could be naturally recast in game theoretic terms. Assume $u^i(a,e^i)$ represents the ordering $\geq (e^i)$. Then Φ is incentive compatible for i in e if and only if e^i is i's best replay to z^i for the payoff function $v^i(s^i,z^i,e^i) = u^i[\Phi(s^i,z^i),e^i]$. Thus, Φ is incentive compatible in e

if and only if e is a Nash equilibrium* of the game with players $i = 1,\dots,n$, strategy sets E^i and payoffs $v^i(s^i,z^i,e^i) = u^i[\Phi(s^i,z^i),e^i]$. A natural generalization of this to an incomplete information game is found in Harsayani [5]. There, strategies are functions $\beta^i : E^i \to E^i$ and a Bayes equilibrium in e given Ψ is a vector of strategies $(\beta^1,\dots,\beta^n)$ such that $\beta^i(e^i)$ solves for each i and all $e^i \varepsilon E^i$ the problem

$$\max_{s^i \varepsilon E^i} \int u^i[\Phi(s^i,\ \beta)^{i(}(e),\ e^i]\ d\ \Psi^i(e^i).$$

Thus, Φ is incentive compatible in e given ψ if and only if the identity map, $\beta^i(e^i) = e^i$ for all i and $e^i \varepsilon E^i$, is a Bayes Equilibrium in e given ψ.

Remark: In Harsayani [5] and other papers, a consistency hypothesis is sometimes used. In particular, it is sometimes assumed that for each i and $e \varepsilon E$, $\psi^i(e^i) = \Pi(\cdot \mid e^i)$ where $\Pi \varepsilon M(E)$ and $\Pi(\cdot \mid e^i)$ is the regular conditional probability measure on z^i given e^i and Π. We will look at restrictions like this below. For now one should only note that the model does not rule out considerations of consistency.

B. Characterizations

We now turn to a collection of results which characterize the incentive compatability of an allocation mechanism under incomplete information in terms of the incentive properties of that mechanism under

---

*It is easy to see that this definition of incentive compatibility is equivalent to the requirement that e^i be a dominant strategy for i in e for all $e \varepsilon E$ and all i.

complete information. As we will see, if the spaces of environments and expectations are broad enough, incentive compatibility under incomplete information obtains if and only if incentive compatibility obtains under complete information.

Theorem 1: An allocation mechanism $\Phi : E \to A$ is incentive compatible in $E \times \Psi$ if and only if Φ is incentive compatible on E.

Proof: (if) If Φ is not incentive compatible on $E \times \Psi$, there is $(e,\psi) \varepsilon E \times \Psi$, some i, and some $s^i \varepsilon E^i$ such that

$$\int u^i[\Phi(s^i,z^i),e^i]\, d\psi^i(e^i) > \int u^i[\Phi(e^i,z^i),e^i]\, d\psi^i(e^i).$$

Therefore, there is an i, $(e^i,z^i) \varepsilon E^i$, and $s^i \varepsilon E^i$ such that

$u^i[\Phi(s^i,z^i), e^i] > u^i[\Phi(e^i,z^i), e^i]$. Thus, Φ is not incentive compatible on E which is a contradiction.

(only if) Suppose Φ is not incentive compatible in $e^* \varepsilon E$. Then there are i and $s^i \varepsilon E^i$ such that $u^i[\Phi(s^i, z^{*i}), e^{*i}] >$ $u^i[\Phi(e^*),e^{*i}]$ where $e^* = (e^{*i},z^{*i})$. Let $\psi \varepsilon \Psi$ be such that $\psi^i(e^i) = \mu^i$ for all e^i where $\mu^i(\{z^{*i}\}) = 1$. Then $\int u^i[\Phi(s^i,z^i),e^{*i}]\, d\psi^i(e^{*i}) =$

$\int u^i[\Phi(s^i,z^i),e^{*i}]\, d\mu^i = u^i[\Phi(e^{*i},z^i), e^{*i}] > u^i[\Phi(e^{*i},z^{*i}),e^{*i}] =$

$\int u^i[\Phi(e^{*i},z^i),e^{*i}]\, d\psi^i\ (e^{*i})$. Thus Φ is not incentive compatible in $E \times \Psi$ which is a contradiction. Q.E.D.

An obvious corollary can be stated which indicates that one does not need incentive compatibility on all of $E \times \Psi$ to ensure incentive compatibility on E. This will be a recurring theme in this paper.

Corollary 1.1: An allocation mechanism $\Phi : E \to A$ is incentive compatible on E if and only if Φ is incentive compatible on $E \times \Psi^*$, where $\psi \varepsilon \Psi^* \subseteq \Psi$ if and only if ψ^i is a constant function on E^i for i = 1, ..., n.

The force of this theorem and its corollary is that if a mechanism is not incentive compatible under complete information then it will also lack incentive compatibility under incomplete information. Thus, the conjecture that the introduction of incomplete information might lead to incentive compatibility, even if it did not obtain under complete information, is shown to be false.

Two objections can be raised at this point which we will consider in sequence. To prove Theorem 1, one uses the obvious fact that if Φ is not incentive compatible for i in e and if i thinks the environment is e (with probability 1) then Φ will still not be incentive compatible for i under "incomplete information." It can be fairly said that, in requiring incentive compatibility for priors concentrated on single environments, we have not retained the spirit behind the introduction of incomplete information and that we should only consider priors which are more diffuse.*

As it turns out, if we are willing to consider only a slightly narrower class of environments and mechanisms, then the result of Theorem 1 still holds even if we restrict priors to be diffuse. To capture the idea of diffuseness, we consider a set $M^+(Z^i) \subseteq M(Z^i)$ of measures on Z^i

---

*Usually a measure is diffuse if $\mu(\{z^i\}) = 0$ for all $z^i \varepsilon Z^i$. Our notion will also require additional properties.

which can be represented by a continuous, positive density function. That is, $\eta \in M^{+}(Z^{i})$ if and only if there is a continuous function $h : Z^{i} \to \mathbb{R}$ such that (a) $h(z^{i}) > 0$ for all $z^{i} \in Z^{i}$ and (b) for any Borel subset $B \subseteq Z^{i}$, $\eta(B) = \int_{B} h^{i}(z^{i})\, dz^{i}$. We will let $\Psi^{+} \subseteq \Psi$ be the set of expectations functions ψ whose range is confined to $M^{+}(Z^{i})$ for each ψ^{i}.

Up to now we have only required that Z^{i} be a Hausdorff topological space. There may be many such Z^{i} for which $M^{+}(Z^{i})$ is empty. Thus, to ensure that $M^{+}(Z^{i}) \neq \emptyset$ we will make the following assumption* when dealing with $M^{+}(Z^{i})$.

Assumption A.1: (a) $A \subseteq \mathbb{R}^{K}$ (K- dimensional Euclidean Space) and $E^{i} \subseteq \mathbb{R}^{M}$ for $i = 1, \ldots, n$ where K and M are finite positive integers.

(b) For all i, either $E^{i} = \mathbb{R}^{M}$ or $0 < \int de^{i} < \infty$.

An example of A and E satisfying A.1 is the class of neo-classical exchange environments with log linear utility functions. There are also many others. Under A.1, $M^{+}(Z^{i}) \neq \emptyset$ since if $E^{i} = \mathbb{R}^{M}$, the multi-variate normal on $\mathbb{R}^{(n-1)M}$ belongs to $M^{+}(Z^{i})$ and if $0 < \int_{E^{i}} de^{i} < \infty$ then the uniform density on Z^{i} is in $M^{+}(Z^{i})$.

The next assumption is used occasionally throughout the rest of this paper.

Assumption A.2: (a) For each $i = 1, \ldots, n$, $u^{i}(a,e^{i})$ is continuous on $A \times E^{i}$.

(b) Φ is continuous on E.

We can now state and prove the following corollary which is used to support the statement that requiring diffuseness does not blunt the impact of Theorem 1.

---

*The assumption A.1 could be weakened at the cost of some confusion.

Corollary 1.2: Under assumptions A.1 and A.2, an allocation mechanism $\Phi : E \to A$ is incentive compatible on $E \times \Psi^{+}$ if and only if Φ is incentive compatible on E.

Proof: (if) Follows from the "if" statement of Theorem 1 since $\Psi^{+} \subseteq \Psi$.

(only if) If Φ is not incentive compatible in e^{*}, there are i and $s^{i} \in E^{i}$ such that

$$(1) \quad u^{i}[\Phi(s^{i}, z^{*i}), e^{*i}] > U^{i}[\Phi(e^{*}), e^{*i}] \text{ where } e^{*} = (e^{*i}, z^{*i}).$$

We consider two cases. Case (1) $\int_{Z^{i}} dz^{i} < \infty$. By assumption A.2, (1) holds on an open neighborhood, N, of z^{*i}. One can choose a continuous function $h:Z^{i} \to \mathbb{R}$ such that $h(z^{i}) > 0$ if $z^{i} \in N$, $h(z^{i}) = 0$ otherwise, and $\int hdz^{i} < \infty$ (by Urysohn's Lemma). Let $h^{*}_{\delta} : Z^{i} \to \mathbb{R}$ be defined by $h^{*}_{\delta}(z^{i}) = h(z^{i}) + \delta$ for all $z^{i} \in Z^{i}$. Let $g_{\delta}(z^{i}) = h^{*}_{\delta}(z^{i})/\delta \cdot \int dz^{i}$, and let $g^{*}(z^{i}) = g_{\delta}(z^{i}) / \int g_{\delta} dz^{i}$. Then if $\delta > 0$, g^{*}_{δ} is continuous, $g^{*}_{\delta}(z^{i}) > 0$ for all $z^{i} \in Z^{i}$ and $\int g^{*}_{\delta}\, dz^{i} = 1$. Let $\eta^{*}_{\delta} \in M^{+}(Z^{i})$, be defined, for all Borel sets $B \subseteq Z^{i}$, by $\eta^{*}_{\delta}(B) = \int_{B} g^{*}_{\delta}(z^{i})\, dz^{i}$, and let $\psi^{*i}_{\delta} \in \Psi^{i+}$ be given by $\psi^{*i}_{\delta}(e^{i}) = \eta^{*}_{\delta}$ for all $e^{i} \in E^{i}$. Now $\int u^{i}[\Phi(s^{i},z^{i}),e^{*i}]\, d\psi^{*i}_{\delta}(e^{*i}) =$ $\int u^{i}[\Phi(s^{i},z^{i}),e^{*i}]\, g^{*}_{\delta}\, dz^{i} = c \cdot \int u^{i}[\Phi(s^{i},z^{i}),e^{*i}] h^{*}_{\delta}\, dz^{i}$, where $c > 0$, which is equal to $\delta \cdot c \cdot \int u^{i}[\Phi(s^{i},z^{i}), e^{*i}]\, dz^{i} +$

$$c \cdot \int_{N} u^{i}[\Phi(s^{i},z^{i}), e^{*i}]\, hdz^{i}.$$

Similarly, $\int u^{i}[\Phi(e^{*i},z^{i}), e^{*i}]\, d\psi^{*i}(e^{*i}) =$

$$\delta \cdot c \cdot \int u^{i}[\Phi\, e^{*i},z^{i}), e^{*i}]\, dz^{i} + c \int_{N} u^{i}[\ (e^{*i},z^{i}), e^{*i}]$$

By (1) and continuity and choice of N,

$$\int_N u^i[\Phi(s^i,z^i),\ e^{*i}]\ hdz^i > \int_N u^i[\Phi(e^{*i},z^i),\ e^{*i}]\ hdz^i.$$

Therefore for δ small enough (since $c < \infty$ and u^i bounded) Φ is not incentive compatible for i in e^* given ψ_δ^{*i}, which is a contradiction.

Case (2): $E^i = \mathbb{R}^M$. Let $f_\delta^h : E^h \to \mathbb{R}$ be the multivariate normal with mean e^{*h} and variance-covariance matrix δI where $\delta > 0$ and I is the identity matrix. Let $g_\delta^*(z^i) = f_\delta^1(e^1) \times \ldots \times f_\delta^{i-1}(e^{i-1}) \times f_\delta^{i+1}(e^{i+1}) \times \ldots \times f_\delta^n(e^n)$. It is easy to show, following case (1), that for δ small enough, Φ is not incentive compatible in e^* given ψ_δ^{*i} (defined as in case (1)) which is a contradiction. Q.E.D.

Remark: While the proof is somewhat tedious, the basic idea is simply to stack enough probability on a small enough neighborhood of z^{*i} in such a way that the prior belongs to $M^+(Z^i)$. Then, with sufficient continuity, even though i considers all z^i possible, he thinks the probability that $z^i \in N$ is close to 1 and, therefore, that he can gain by misrepresenting e^{*i} as s^i. One must thus conclude that the introduction of diffuse incomplete information alters none of the incentive properties of continuous mechanisms in environments with continuous preferences.

Remark: As before, Corollary 1.2 remains valid if Ψ^+ is replaced with $\Psi^+ \cap \Psi^*$, the set of functions which take a constant value in $M^+(Z^i)$ for each i. As we will see below (Corollary 1.3) the set Ψ^+ can be restricted even further.

The second objection which can be raised against Theorem 1 is that no consistency is required of the agents' expectations. It is

usually argued that consistency is a long-run phenomenon (after learning has occured). Under this interpretation, consistent agents have more information than inconsistent agents and, therefore, a lack of incentive compatibility should be more prevalent. Whether this is true or not, consistency is a natural restriction in the context of the Harsayani game model and should be considered.

We will say an expectations function $\psi \in \Psi$ is consistent if there is $\Pi \in M(E)$ such that, for i = 1, ..., N, $\psi^i(e^i) = \Pi(\cdot|e^i)$ where $\Pi(\cdot|e^i)$ is the regular conditional probability measure over Z^i given e^i and Π. We let R_E be the set of consistent $\psi \in \Psi$.

Corollary 1.3: Under assumptions A.1 and A.2, an allocation mechanism $\Phi: E \to A$ is incentive compatible on* $E \times (R_E \cap \Psi^+)$ if and only if Φ is incentive compatible on E.

Proof: (if) Since $R_E \cap \Psi^+ \subseteq \Psi$, this follows directly from Theorem 1.

(only if) Suppose Φ is not incentive compatible in e^*. Let ψ_δ^{*i} be as constructed in the proof of corollary 1.2. Φ is not incentive compatible in e^{*i} given ψ_δ^{*i} for δ small enough, $\delta > 0$. It remains to show that $\exists\ \Pi \in M(E)$ such that $\psi_\delta^{*i}(e^i) = \Pi(\cdot|e^i)$ for all $e^i \in E^i$ and such that $\Pi(\cdot|e^i) \in M^+(Z^h)$ for all $e^h \in E^h$. Let $g^i: E^i \to \mathbb{R}$ be any continuous, positive density function on E^i. Let $g_\delta^*: Z^i \to \mathbb{R}$ be as in Corollary 1.2. Finally let $G(e) = g^i(e^i) \cdot g_\delta^*(z^i)$ for all $e \in E$. Define $\Pi \in M^+(e)$ by, for each Borel subset $B \subseteq E$, $\Pi(E) = \int_B G(e)de$. Let $\psi^h(e^h) = \Pi(\cdot|e^h)$. Then, $\psi \in R_E \cap \Psi^+$ and Φ is not incentive compatible at e^* given ψ which is a contradiction. Q.E.D.

---

*We chose $R_E \cap \Psi^+$ instead of R_E since if the corollary is valid on $R_E \cap \Psi^+$, then it is valid on R_E and since $\psi \in R_E \cap \Psi^+$ if and only if beliefs are consistent in the sense of [1, p. 24].

Remark: Corollary 1.3 remains valid if $R_E \cap \Psi^+$ is replaced with* $R_E \cap \Psi^+ \cap \Psi^*$. This is shown directly from the proof if $E^h = \mathbb{R}^m$ for all h since then G(e) is composed of independently distributed multivariate normals. And if the e^h are independently distributed for some continuous positive density G(e), then the ψ derived will belong to $R_E \cap \Psi^+ \cap \Psi^*$. If $\int dz^i < \infty$, as in the case (1) of Corollary 1.2, then a different g^*_δ must be produced which has the property that the e^h, $(h \neq i)$, are independently distributed. This can be done by decomposing the neighborhood N into neighborhoods N^h for h = 1, ..., n, $h \neq i$ such that $N^1 \times \ldots \times N^{i-1} \times \ldots \times N^n \subseteq N$. Then one follows the same procedure as in corollary 1.2 to construct g^{*h}_δ for h = 1, ..., n, and $n \neq i$. Let $G(e) = g^{*1}_\delta(e^1) \times \ldots \times g^{*n}_\delta(e^n)$ where $g^{*i}_\delta(e^i)$ is essentially arbitrary. Then the derived ψ belongs to $R_E \cap \Psi^+ \cap \Psi^*$ and Φ is not incentive compatible in e^* given ψ.

Remark: Given the validity of corollary 1.3 for

$R_E \cap \Psi^+ \cap \Psi^* \equiv I \subseteq \Psi$, the validity of Corollary 1.1, 1.2, and 1.3 follows from Theorem 1 and the fact that $I \subseteq \Psi^*$, $I \subseteq \Psi^+$, and $I \subseteq R_E \cap \Psi^+$. In fact, one can easily show that if $I \subseteq B \subseteq \Psi^+$, then under A.1 and A.2, Φ is incentive compatible on E if and only if Φ is incentive compatible on E × B. There are undoubtedly smaller sets than I which have this property. It is easy to prove the following:

---

*Requiring that $\psi \in R_E \cap \Psi^+ \cap \Psi^*$ is equivalent to the requirement that there are, for i = 1, ..., n, continuous positive densities, g^i, on E^i such that $\psi^i(e^i)(A) =$

$$\int_A g^1(e^1) \times \ldots \times g^{i-1}(e^{i-1}) \times g^{i+1}(e^{i+1}) \times \ldots \times g^n(e^n)dz^i.$$

Compare this to the concept of "independence" defined in [1, p. 24].

Lemma 1: Suppose Φ is incentive compatible on E if and only if Φ is incentive compatible on $E \times Q \subseteq E \times \Psi$, and suppose that $Q \subseteq B \subseteq \Psi$. Then Φ is incentive compatible on E if and only if Φ is incentive compatible on E x B.

The converse is not necessarily true.

Remark: One subset of environments which is of particular interest to social choice theorists occurs when the sets A and E^i are restricted to be finite.* E^i is usually the set of all profiles (subsets of A x A) on A and is thus finite if A is. Let E_F be this set of environments. Since E_F and A are finite, the only Hausdorf topology is the discrete topology (all subsets are both open and closed). Thus all functions from E to A are continuous and all functions from $A \times E^i$ to $\mathbb{R}$ are continuous. Thus if A.0 holds then A.1(a) and A.2 also do. Finally, since E^i can be parameterized such that $E^i \subseteq \mathbb{R}$, $\int de^i$ = the cardinality of E^i which is finite so A.1(b) holds. Thus, Theorem 1 and all its corollaries apply when E is replaced by E_F. In fact, the proofs are easier since $\eta \in M^+(Z^i)$ if and only if there are positive numbers P_z for each z such that $\eta(B) = \sum_{z \in B} P_z$, since under the discrete topology, any assignment of probabilities to the finite set Z^i is continuous. For example, let $\eta(B)$ = cardinality of B/cardinality of Z^i. Then $\eta \in M^+(Z^i)$.

One can summarize the results of this section in a single sentence. In spite of restrictions with respect to diffuseness,

---

*This is the case for which Gibbard [2] and Satterthwaite [14] show that Φ is incentive compatible in E if and only if Φ is dictatorial.

independence, or finite alternatives, if the set of possible expectations is broad enough and preferences are continuous, *a continuous allocation mechanism is incentive compatible under incomplete information if and only if it is incentive compatible under complete information.* Since most of the mechanisms we know of are not incentive compatible under complete information, we are left with the unpleasant fact that *most of the mechanisms we know of will not be incentive compatible under incomplete information.* *

C. Generic Properties

It has been demonstrated that an allocation mechanism which lacks incentive compatibility under complete information also lacks incentive compatibility under incomplete information for some prior beliefs. One might next ask for how many priors is this true. If there are only a few, then there is some hope that incentive compatibility could be rescued with the introduction of incomplete information. As we will see, however, the set of priors leading to a lack of incentive compatibility is large. In particular, if there is just one environment (e^i, z^i) in which Φ is not incentive compatible for i then Φ is not incentive compatible for i on an open dense subset of prior beliefs given e^i.

We first show that, for continuous preferences, mechanisms, and expectations functions, incentive compatibility is a closed property.

*This appears to be at odds with the results in [1]. For a discussion of this "paradox", see the last section of this paper.

Theorem 2: Under assumption A.2,

(a) an allocation mechanism $\Phi:E\to A$ is incentive compatible on a closed subset (possibly empty) of E,

(b) an allocation mechanism $\Phi:E\to A$ is incentive compatible on a (relatively*) closed subset of $E \times \Psi^c$, where $\psi \in \Psi^c$ if and only if ψ is a continuous function.

Proof: (a) Suppose Φ is not incentive compatible for some i in $(e^i,z^i) \in E$. Then there is $s^i \in E^i$ such that $u^i\,[\Phi(s^i,z^i),\, e^i] > u^i\,[\Phi(e^i,z^i),\, e^i]$. But u^i and Φ are continuous. Therefore, the inequality holds on an open neighborhood of (e^i,z^i). The conclusion follows easily.

(b) Suppose Φ is not incentive compatible in e^* given ψ for some $(e^*, \psi) \in E \times \Psi^c$. Then there are i and $s^i \in E^i$ such that

$$\int u^i[\Phi(s^i,z^i),\, e^{*i}]\, d\psi^i(e^{*i}) > \int u^i[\Phi(e^{*i},z^i),\, e^{*i}]\, d\psi^i(e^{*i}).$$

Now u^i is continuous and bounded, Φ is continuous, and ψ^i is continuous. Further, in the topology of weak convergence on $M(Z^i)$, $\eta \to \overline{\eta}$ if and only if $\int f d\eta \to \int f d\overline{\eta}$ for all continuous bounded functionals f. Let $f(e^{*i},z^i) = u^i[\Phi(s^i,z^i),e^{*i}] - u^i[(e^{*i},z^i),e^{*i}]$. It follows easily that there is a relatively open neighborhood N of (e^*,ψ) where $N \subseteq E \times \Psi^c$ such that if $(e,\psi^*) \in N$ then $\int f(e^i,z^i)\, d\psi^{*i}(e^i) > 0$.

*B is a relatively closed set of $A \subseteq E \times \Psi$ if and only if $B = A \cap C$ where C is a closed subset of $E \times \Psi$.

The conclusion follows since Φ is not incentive compatible on an open subset of $E \times \Psi^c$. Q.E.D.

Corollary 2.1: Under assumption A.2, Theorem 2(a) is valid when any subset $B \subseteq E$ replaces E and Theorem 2(b) is valid when any subset $B \subseteq E \times \Psi^c$ replaces $E \times \Psi^c$.

Proof: Follows from Theorem 2 and the definition of relative topology. Q.E.D.

Remark: Since the set of constant expectation functions Ψ^* is a subset of Ψ^c, Theorem 2 applies on subsets of $E \times \Psi^*$. If ψ does not belong to Ψ^c (i.e., ψ^i is not continuous) then small perturbations in e^i could cause the inequality, $\int f(e^i,z^i)\, d\psi^i\,(e^i)$, to switch. Thus Theorem 2 will not necessarily hold on subsets of Ψ not contained in Ψ^c. Examples of such broader sets are R_E and Ψ^+. However, the theorem will hold (by the corollary) on $\Psi^c \cap R_E$ and $\Psi^c \cap \Psi^+$.

To establish conditions under which incentive compatibility with incomplete information is a nowhere dense property, we will need to restrict further the class of environments and mechanisms under consideration.

Assumption A.3: (a) $A \subseteq \mathbb{R}^K$ and, for each i = 1, ..., n, $E^i \subseteq \mathbb{R}^M$ where K and M are finite positive integers,

(b) Φ has continuous first derivatives in e on E and, for each i = 1, ..., n, u^i has continuous first derivatives in $a \in A$ given e^i on $A \times E^i$,

(c) for each i = 1, ..., n and each $(e^i,z^i) \in E^i \times Z^i$, $u^i[\Phi(e^i,z^i), e^i] \geq u^i[\Phi(s^i,z^i),e^i]$ for all $s^i \in E^i$ if and only if

$\nabla u^i \cdot \nabla_i \Phi(e^i,z^i) = 0$ where $\nabla u^i \equiv (\partial u^i/\partial a_1, \ldots, \partial u^i/\partial a_K)$, $\nabla_i \Phi \equiv (\partial\Phi/\partial e^i_1, \ldots, \partial\Phi/\partial e^i_M)$, and $\nabla u^i \cdot \nabla_i \Phi(e^i,z^i)$ is the vector of partial derivatives of $u^i[\Phi(s^i,z^i), e^i]$ with respect to $(s^i_1,\ldots,s^i_M)$ evaluated at $s^i = e^i$,

(d) for each i = 1, ..., n, for each $e^i \in E^i$, and for each $\eta \in M(Z^i)$, $|\, \nabla u^i \cdot \nabla_i \Phi(e^i,z^i) \,|$ is bounded by an integrable function of z^i.

An example of a class of environments (neoclassical exchange with Cobb-Douglas utility) and a mechanism (the competitive process) which satisfy A.3 is given in [9].

To characterize situations in which incentive compatibility is a nowhere dense property we will need to define one more set. Let $E^*_\Phi \equiv \{e \in E \mid$ for i = 1, ..., n, Φ is incentive compatible for i on $\{e^i\} \times Z^i\}$. We note in passing that if $\hat{E}_\Phi$ is the set of e such that Φ is incentive compatible in e, then $E^*_\Phi \subseteq \hat{E}_\Phi$ and the inclusion is usually strict. In fact, E^*_Φ may be empty even though $\hat{E}_\Phi$ is not. Also $\hat{E}_\Phi$ can be a dense subset of E while E^*_Φ is nowhere dense.

Theorem 3: Under assumption A.3, an allocation mechanism $\Phi: E \to A$ is incentive compatible only on a nowhere dense subset of $E \times \Psi$ if and only if E^*_Φ is a nowhere dense subset of E.

Proof: (only if) Let $e^* \in E^*$. Then by referring to the proof of the "if" statement of Theorem 1 it follows that Φ is incentive compatible in (e^*,Φ) for all $\psi \in \Psi$. Let $\psi^* \in \Psi$. By hypothesis, for any neighborhood $N \subseteq E \times \Psi$ of (e^*, ψ^*) there is $(e', \psi') \in N$ such that Φ is not incentive compatible in e' given ψ'. Thus, again as in the proof of the "if" statement of Theorem 1, it follows that there is i, $s^i \varepsilon E^i$, $\bar{z}^i \varepsilon Z^i$ such that $u^i[\Phi(s^i,\bar{z}^i),e'^i] > u^i[\Phi(e'^i,\bar{z}^i),e'^i]$. Thus, $e' \notin E^*$. Thus, for every neighborhood $N^* \subseteq E$ of $e \varepsilon E$ there is $e' \varepsilon N^*$ such that $e' \notin E^*$. This establishes the proposition.

(if) Suppose Φ is incentive compatible for e given ψ where $(e, \psi) \varepsilon E \times \Psi$. By hypothesis, for any neighborhood $N \subseteq E \times \Psi$ of (e,ψ) there is $\hat{e} \varepsilon E$ such that $\hat{e} \notin E^*$ and $(e,\psi) \varepsilon N$. If Φ is not incentive compatible in $\hat{e}$ given ψ, we are through. If Φ is incentive in $\hat{e}$ given ψ, we can perturb ψ slightly and destroy the incentive compatibility. To see this, we note first that since $\hat{e} \notin E^*$, there is i, and $\bar{z}^i \varepsilon Z^i$ such that $\nabla u^i \nabla_i \Phi(\hat{e}^i,z^i) \neq 0$ [by A.3(c)]. Further, if Φ is incentive compatible in $\hat{e}$ given ψ, then again by A.3 it is true that $\int \nabla u^i \nabla_i \Phi(e^i,z^i) d\Psi^i(\hat{e}^i) = 0$. Consider the new expectations function ψ^i_δ defined for each $e^i \varepsilon E^i$, each Borel set $B \subseteq Z^i$, and each $s \geq 0$ by,

$$\psi^i_\delta(e^i)(B) = (1-\delta)\ \psi^i(e^i)(B) \qquad \text{if } \bar{z}^i \notin B$$
$$(1-\delta)\ \psi^i(e^i)(B) + \delta \quad \text{if } \bar{z}^i \varepsilon B$$

Clearly $\delta \to 0$ implies $\psi^i_\delta \to \psi^i$. Further,

$$\int \nabla u^i\ \nabla_i \Phi(e^i,z^i) d\ \psi^i_\delta(\hat{e}^i) = (1-\delta)\ \nabla u^i\ \nabla_i \Phi(\hat{e}^i,z^i) d\psi^i(e^i)$$
$$+\ \delta \cdot \nabla u^i \cdot \nabla_i \Phi(\hat{e}^i,\bar{z}^i) = \delta\ \nabla u^i \nabla_i\ \Phi(\hat{e}^i,\bar{z}^i) \neq 0 \quad \text{for all } \delta > 0.$$

Thus for any neighborhood N of $(\hat{e},\psi)$ we can choose δ small enough such that $(\hat{e},\psi_\delta) \equiv (\hat{e}, \psi^1,\ldots, \psi^{i-1}, \psi^i_\delta, \psi^{i+1},\ldots, \psi^n) \varepsilon N$ and such

that Φ is not incentive compatible in $\hat{e}$ given ψ. The desired conclusion follows. Q.E.D.

Corollary 3.1: Under assumption A.3, E^*_Φ is a nowhere dense subset of E given the allocation mechanism $\Phi: E \to A$ if and only if Φ is incentive compatible only on a closed, nowhere dense subset of $E \times \Psi^C$.

Proof: Closedness follows from Theorem 2. Nowhere denseness follows as in Theorem 3 by noting that if $\psi^i \varepsilon \Psi^{iC}$ then $\psi^i_\delta \varepsilon \Psi^{iC}$ since the perturbation employed does not destroy continuity. Q.E.D.

Remark: One concludes from these results that, with enough differentiability, if the lack of incentive compatibility is generic for Φ in E then it is also generic in $E \times \Psi^C$. Also even if lack of incentive compatibility is not generic in E, if E^* is nowhere dense, then the set of expectations and environments on which Φ lacks incentive compatibility is still large (the set contains an open dense subset of $E \times \Psi^C$.

As before, it is desirable to inquire whether Theorem 3 remains valid if we further restrict the class of expectations functions. It is easy to verify that if B is a subset of Ψ with a non-empty interior and if Φ is not incentive compatible on a set $A \subseteq E \times \Psi$ which is dense in $E \times \Psi$ then Φ is not incentive compatible on $C \equiv A \cap (E \times B) \subseteq E \times B$ where C is dense in $(E \times B)$. Unfortunately most of the interesting sets of expectations may not have non-empty interiors. For example, suppose $\psi^i \varepsilon \Psi^{+1}$. That is, $\psi^i : E^i \to M^+(Z^i)$. Let ψ^i_δ be defined as in the proof of Theorem 3. It is true that $\psi^i_\delta \notin \Psi^{+i}$ for all $\delta > 0$. Therefore, Ψ^{+i} does not have a non-empty interior in Ψ. Luckily, a non-empty interior is not a necessary condition. Thus we have the following corollary.

Corollary 3.2: Theorem 3 remains valid if Ψ is replaced by Ψ^* or Ψ^+ or R_E or Ψ^C or by any combination of intersections of these sets. In particular, Theorem 3 remains valid if Ψ is replaced with $R_E \cap \Psi^+ \cap \Psi^*$ (the set of consistent beliefs satisfying an independence condition).

Proof: First, it is easy to verify that the proof of the "only if" part of Theorem 3 survives if Ψ is replaced by any non-empty subset. Second, it is also easy to verify that the proof of the "if" part of Theorem 3 survives if Ψ is replaced by any non-empty subset B for which the following property is true: if there is i and $e^* \varepsilon E$ such that (i) $\nabla u^i \cdot \nabla_i \Phi(e^{*i}, z^{*i}) \neq 0$ and (ii) if $\psi \varepsilon B$ and $\int \nabla u^i \nabla_i \Phi(e^{*i}, z^i) d\psi^i(e^{*i}) = 0$ then for any neighborhood N of ψ in B, there is $\overline{\psi}^i \varepsilon N$ such that $\int \nabla u^i \nabla_i \Phi(e^{*i}, z^i) d\overline{\psi}^i(e^{*i}) \neq 0$.

Let $\psi \varepsilon \Psi$ satisfy the condition that there is i and $e^* \varepsilon E$ such that (i) and (ii) are true. By continuity (i) holds on an open neighborhood $N \subseteq Z^i$ of z^{*i}. Let $h: Z^i \rightarrow \mathbb{R}$ be any continuous non-negative function such that $\int_N h(z^i) dz^i = 1$. (These are easy to construct since $Z^i \subseteq \mathbb{R}^{(n-1)M}$.) Let $\eta^* \varepsilon M(Z^i)$ be defined by, for all Borel sets $B \subseteq Z^i$, $\eta^*(B) = \int_B h(z^i) dz^i$. Finally let $\psi^i_\lambda(e^i)(B) = (1 - \lambda)\psi^i(e^i)(B) + \lambda\eta^*(B)$ when B is a Borel subset of Z^i. The following properties are true, whenever $\lambda \varepsilon (0,1)$: (a) if $\psi^i \varepsilon \Psi^{*i}$, then $\psi^i_\lambda \varepsilon \Psi^{*i}$, (b) if $\psi^i \varepsilon \Psi^{ci}$, then $\psi^i_\lambda \varepsilon \Psi^{ci}$, then $\psi^i_\lambda \varepsilon \Psi^{ci}$, and (c) if $\psi^i \varepsilon \Psi^{+i}$, then $\psi^i_\lambda \varepsilon \Psi^{+i}$. It is also true that, if (i) holds, and $\lambda \varepsilon (0,1)$, then

$\int u^i \nabla_i \Phi(e^{*i}, z^i) d\psi^i_\lambda(e^{*i}) \neq 0$. Therefore, the corollary is valid for Ψ^* or Ψ^+ or Ψ^c or any combination of intersections of these.

Finally, suppose $\psi \varepsilon R_E$. Then there is $\Pi \varepsilon M(E)$ such that $\psi^i(e^i) = \Pi(\cdot|e^i)$ for all i and all $e^i \varepsilon E^i$. By continuity (i) holds on an open neighborhood N of e^*. Let $h: E \rightarrow \mathbb{R}$ be a continuous non-negative function such that $\int_N h de = 1$, and $\int h(e^{*i}, z^i) dz^i > 0$, and let $\eta(B) = \int_B h(e) de$ for all Borel sets $B \subseteq E$. Let $\Pi_\lambda = (1 - \lambda)\Pi(B) + \lambda\eta^*(B)$ for Borel sets $B \subseteq E$. Finally let $\psi^i_\lambda(e^i) = \Pi_\lambda(\cdot|e^i)$. Then $\psi_\lambda \varepsilon R_E$. Further, given (i), $\int \nabla u^i \nabla_i \Phi(e^{*i}, z^i) d\psi^i_\lambda(e^{*i}) \neq 0$ for $\lambda \varepsilon (0,1)$. Thus, the corollary is valid for R_E. Finally, if $\psi \varepsilon \Psi^*$ or Ψ^+ or Ψ^c then so does ψ_λ. Thus, the corollary is proven. Q.E.D.

Remark: The subset of E, with E^i and A finite, that is of interest to Social Choice theorists is not covered by Theorem 3 or its corollaries, since A.3 will not hold for this class of environments. In particular, the derivatives ∇u^i and $\nabla_i \Phi$ are not even defined.

One can summarize the results of this section in a single sentence. Since for most mechanisms we know of E^*_Φ is nowhere dense, we must conclude that, in spite of restrictions with respect to diffuseness or independence of expectations, *for most differentiable mechanisms and environments, incentive compatibility will usually not obtain even if information is incomplete.*

3. Concluding Remarks

In this section we explore some of the implications of our results for specific environments and mechanisms.

A. Exchange Economies

The theorem of Hurwicz [7], that there is a neo-classical exchange economy such that no allocation mechanism $\Phi: E \rightarrow A$ with the properties that $\Phi(e)$ is Pareto-optimal in e and $\Phi(e) \geq (e^i)\ a_o$ (where a_o is the no trade allocation) is incentive compatible implies, when combined with theorem 1, that no such mechanism is incentive compatible on neoclassical economies under incomplete information. In particular, the competitive mechanism is not incentive compatible on the class of neoclassical economies under either complete or incomplete information.

The work of Mantel [10] and others can be used to show that given the preferences of one consumer and a price vector, p, there is an economy with that consumer such that p is the equilibrium price. Thus, one can for each $e^i \varepsilon E^i_N$ (the set of neoclassical characteristics) find $z^i \varepsilon Z^i$ (possibly varying n, the number of consumers) such that the initial endowment is not Pareto-optimal. It is a short step to show that $E^*_\Phi = \emptyset$ if Φ is the competitive process. Thus, the competitive process is not incentive compatible on an open dense subset of, say, $R_E \cap \Psi^+ \cap \Psi^*$ or of $\Psi^+ \cap \Psi^c$ or of Ψ^*. A similar conclusion would follow, via the construction of Hurwicz [7], for any decentralized mechanism which selects Pareto-efficient outcomes and gives a "no-trade" option. Thus, under incomplete information, the lack of incentive compatibility is generic for all such mechanisms.

B. Public Goods

Roberts [13] extends the results of Hurwicz to economies with public goods. It follows, as above, that for any mechanism which selects

Pareto-efficient allocations and which allows a "no-trade" option, the lack of incentive compatibility is generic under incomplete information. Green-Laffont [3], Hurwicz [8], and Walker [15] have dispensed with the no-trade option and have shown that Pareto-efficient mechanisms cannot be incentive compatible on neo-classical economies with public goods. Walker [15] has shown the lack of incentive compatibility to be generic on a subset of those economies. Thus, for at least a subset of neo-classical economies with public goods, a lack of incentive compatibility is generic for all Pareto-efficient mechanisms under either complete or incomplete information.

C. A "Paradox"

The last sentence of the previous section seems to be in direct opposition to the remarkable result of d'Aspremont and Gerard-Varet [1] which states that if a mutual consistency requirement on prior beliefs is satisfied, then there exist Pareto-efficient mechanisms which are incentive compatible under incomplete information on the class of neoclassical public goods economies with transferable utility. Mutual consistency can be interpreted in the context of our model as requiring that $\psi \varepsilon R_E \cap \Psi^+$.

This apparent paradox is resolved by noting that their definition of a mechanism allows Φ to be a function of the probability measure $\Pi \varepsilon M(E)$ where $\psi^i(e^i) = \Pi(|e^i)$ for all i and e^i (when $\psi \varepsilon R_E$). Let $R \subseteq M(E)$ be the set of Π related to some $\psi \varepsilon R_E \cap \Psi^+$. Then their mechanisms are functions* $\Phi: E \times R \rightarrow A$.

---

*A similar class of mechanisms is considered by Meyerson [11] to resolve a bargaining problem.

Returning to corollary 1.3 in which it is shown that Φ is incentive compatible on $R_E \cap \Psi^+$ if and only if on E, one sees that this is proven by finding $\Pi^* \varepsilon R$ such that $\int \nabla u^i \cdot \nabla_i \Phi(e^i, z^i) g^*(e) dz^i \neq 0$ for some i and e^i where g^* is the continuous density representing Π^*. This is done by showing that if $\int \nabla u^i \nabla_i \Phi g^* dz^i = 0$ then one can perturb Π^* to Π_δ such that $\int \nabla u^i \nabla_i \Phi g_\delta dz^i \neq 0$. Some brief thought will convince one that the ability to produce the appropriate perturbation relies on the fact that $u^i[\Phi(e), e^i]$ is invariant with changes in Π or g. For the broader model, where Φ depends on Π we must consider (if defined):

$$(2) \quad \int \nabla u^i [(\nabla_i \Phi) g(s^i, z^i) + (\nabla_g \Phi) \nabla_i g(s^i, z^i)] \, dz^i.$$

Thus, even though $\int \nabla u^i \cdot \nabla_i \Phi \cdot g_\delta dz^i \neq 0$ for some $\delta > 0$, it may be true that this effect is just balanced by $\int \nabla u^i \cdot \nabla g \Phi \cdot \nabla_i g dz^i$. In fact, Theorem 3.3 in [1] shows that there are Φ for which this is true. Thus, for all $\Pi \varepsilon R$, (2) will be zero when evaluated at $s^i = e^i$ and Φ will be immune from manipulations of the form $s^i \neq e^i$. However, it follows from the theorems in the previous sections of this paper, since Π depends on ψ^i and Φ depends on Π, that Φ is not immune to "manipulations" in Π.

I originally concluded in [9] that the only way to avoid misrepresentations of priors in the d'Aspremont-Gerard-Verat model was to assume that each agent knew all other agents' priors but that this assumption violated the spirit of incomplete information and the concept of privacy preserving mechanisms.*

I am no longer convinced that this is a completely valid view but I still have some questions. Since e^i summarizes i's true beliefs (as well as tastes, etc.) one might claim that misrepresentations in prior beliefs can occur only through misrepresentations of e^i and not of Π. Thus, given Π the theorem of d'Aspremont and Gerard-Verat could be interpreted as establishing that, for their mechanism, misrepresentations of beliefs and preferences are unprofitable for all agents. However, one remaining unanswered question is: how is Π "known" by all agents, including the "center" and why can't Π be manipulated? Two justifications of consistency are presented at the end of [1], but this question is never really confronted.

A generous interpretation of the origin of Π, which can be found in Harsayani [5], is that it incorporates all the relevant information and beliefs of a well-informed outside observer† and that each agent is such an observer prior to his discovery of his special information e^i.

Thus in a real sense, each i is assumed to know $E \equiv E^1 \times \ldots \times E^n$ and $\Pi \varepsilon M(E)$ and to agree on this knowledge. If this is the case, then

---

*As presented by Hurwicz in [6] and [7], a mechanism is privacy preserving if the outcome depends solely on the reported information of the agents (a non-parametric outcome).

†It is not at all clear to me how one places the beliefs of an outside observer on the same foundations as the beliefs of a decision maker, as has been done by Savage and others.

Φ can depend on Π and Π cannot be manipulated.

On the other hand, if agreement on Π is arrived at through past experience or some other form of communication then Π can be manipulated since agents can gain by acting as if their beliefs are ψ^{*i} instead of ψ^i whenever $\int u^i[\Phi(e,\Pi(\psi^*)),e^i]\, d\psi^i(e^i) >$ $\int u^i[\Phi(e,\Pi(\psi)), e^i]\, d\psi^i(e^i)$, where $\Pi(\psi)$ is the element of $M(E)$ such that $\psi^i(e^i) = \Pi(\cdot|e^i)$ for all i and e.

In this case the incentive compatibility issue simply arises at an earlier stage of the allocation process. It remains an open question, in my mind, whether there is a justification for universal agreement on Π which does not simply assume away the potentials for manipulation implicit in any form of discovery through communication.

REFERENCES

1. d'Aspremont, C. and L. Gerard-Varet. "Incentives and Incomplete Information," CORE Discussion Paper 7705, Universite Catholique de Louvain, 1977.

2. Gibbard, A. "Manipulation of Voting Schemes: A General Result," Econometrica 41 (1973): 587-601.

3. Green, J. and J. J. Laffont. "Revelation des Preferences pour les Biens Publics", D. P. Ecole Polytechnique, No. A 1250176; to appear in Cahiers du Seminaire d'Econometrie C.N.R.S., Paris.

4. Groves, T. and M. Loeb. "Incentives and Public Inputs," Journal of Public Economics, 4 (1975): 211-226.

5. Harsayani, J. "Games with Incomplete Information Played by 'Bayesian' Players, Parts I-III," Management Science 14 (1967-68): 158-182, 320-334, 486-502.

6. Hurwicz, L. "Optimality and Informational Efficiency in Resource Allocation Processes" in Mathematical Methods in Social Sciences, K. J. Arrow, S. Karlin and P. Suppes, eds., Stanford University, 1960: 27-46.

7. Hurwicz, L. "On Informationally Decentralized Systems," in Decision and Organization, R. Radner and C. B. McQuire, eds., North-Holland: 1972.

8. __________. "On the Existence of Allocation Systems Whose Manipulative Nash Equilibria are Pareto-optimal." Unpublished paper presented at the Third World Congress of the Econometric Society, Toronto, August 1975.

9. Ledyard, J. "The Incentives for Price-Taking Behavior when Consumers Have Incomplete Information," D. P. No. 227, The Center for Mathematical Studies in Economics and Management Science, Northwestern University.

10. Mantel, R. "Homothetic Preferences and Community Excess Demand Functions," Journal of Economic Theory 12 (1976): 197-201.

11. Meyerson, R. "Incentive Compatibility and the Bargaining Problem," D. P. No. 284, The Center for Mathematical Studies in Economics and Management Science, Northwestern University, 1977.

12. Reiter, S. "Informational Efficiency of Iterative Processes and the Size of the Message Space," Journal of Economic Theory 8 (1974): 193-205.

13. Roberts, J. "Incentives and Planning Procedures for the Provision of Public Goods," to appear in Review of Economic Studies, 1977.

14. Satterthwaite, M. "Strategy Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," Journal of Economic Theory 10 (1975): 187-217.

15. Walker, M. "On the Impossibility of a Dominant-Strategy Mechanism to Optimally Decide Public Questions" W.P. No. 160, Department of Economics, State University of New York at Stony Brook, 1976, to appear in Econometrica.